

J U N E 2 0 2 0

REPORT TO THE CONGRESS

**Medicare and the
Health Care
Delivery System**

MEDPAC Medicare
Payment Advisory
Commission

PFE000193



The Medicare Payment Advisory Commission (MedPAC) is an independent congressional agency established by the Balanced Budget Act of 1997 (P.L. 105–33) to advise the U.S. Congress on issues affecting the Medicare program. In addition to advising the Congress on payments to health plans participating in the Medicare Advantage program and providers in Medicare’s traditional fee-for-service program, MedPAC is also tasked with analyzing access to care, quality of care, and other issues affecting Medicare.

The Commission’s 17 members bring diverse expertise in the financing and delivery of health care services. Commissioners are appointed to three-year terms (subject to renewal) by the Comptroller General and serve part time. Appointments are staggered; the terms of five or six Commissioners expire each year. The Commission is supported by an executive director and a staff of analysts, who typically have backgrounds in economics, health policy, and public health.

MedPAC meets publicly to discuss policy issues and formulate its recommendations to the Congress. In the course of these meetings, Commissioners consider the results of staff research, presentations by policy experts, and comments from interested parties. (Meeting transcripts are available at www.medpac.gov.) Commission members and staff also seek input on Medicare issues through frequent meetings with individuals interested in the program, including staff from congressional committees and the Centers for Medicare & Medicaid Services (CMS), health care researchers, health care providers, and beneficiary advocates.

Two reports—issued in March and June each year—are the primary outlets for Commission recommendations. In addition to annual reports and occasional reports on subjects requested by the Congress, MedPAC advises the Congress through other avenues, including comments on reports and proposed regulations issued by the Secretary of the Department of Health and Human Services, testimony, and briefings for congressional staff.

J U N E 2 0 2 0

REPORT TO THE CONGRESS

**Medicare and the
Health Care
Delivery System**

MEDPAC Medicare
Payment Advisory
Commission

425 I Street, NW • Suite 701 • Washington, DC 20001
(202) 220-3700 • Fax: (202) 220-3759 • www.medpac.gov

PFE000195



425 I Street, NW • Suite 701
Washington, DC 20001
202-220-3700 • Fax: 202-220-3759
www.medpac.gov

Francis J. Crosson, M.D., Chairman
Paul B. Ginsburg, Ph.D., Vice Chairman
James E. Mathews, Ph.D., Executive Director

June 15, 2020

The Honorable Michael R. Pence
President of the Senate
U.S. Capitol
Washington, DC 20510

The Honorable Nancy Pelosi
Speaker of the House
U.S. House of Representatives
U.S. Capitol
Room H-232
Washington, DC 20515

Dear Mr. President and Madam Speaker:

I am pleased to submit the Medicare Payment Advisory Commission's June 2020 *Report to the Congress: Medicare and the Health Care Delivery System*. This report fulfills the Commission's legislative mandate to evaluate Medicare payment issues and to make recommendations to the Congress.

In the seven chapters in this report, we consider:

- realizing the promise of value-based payment in Medicare, an agenda for change.
- challenges in maintaining and increasing savings from accountable care organizations.
- replacing the Medicare Advantage quality bonus program.
- the impact of changes in the 21st Century Cures Act to risk adjustment for Medicare Advantage enrollees, a mandated report.
- realigning incentives in Medicare Part D.
- separately payable drugs in the hospital outpatient prospective payment system.
- improving Medicare's end-stage renal disease prospective payment system.

In particular, I wish to draw your attention to Chapter 1, which is the result of a year-long Commission discussion about the future of the Medicare program. The Commission believes that unless substantial changes are made to the way Medicare pays for services and to how beneficiary care is organized and delivered, the cost of the Medicare program will remain on an unsustainable trajectory. The Commission

Page 2

asserts that the use of fee-for-service payment for Medicare services should be replaced, over time and to the degree feasible, by payment to accountable systems of care that have incentives to:

- provide preventive services and early disease detection.
- improve the quality and beneficiary experience of care.
- avoid delivering unnecessary or inappropriate services.
- control the costs of providing necessary services in the most appropriate care setting.
- deliver chronic care services through care coordination among providers.
- coordinate both the medical and nonmedical needs of beneficiaries.
- enhance the use of technologies that improve quality and reduce program costs.

Under an improved Medicare program, most beneficiaries would be able to opt to receive their care through accountable entities. Medicare could design incentives that encourage beneficiaries to choose one of these entities and give providers incentives to participate in them.

The Commission well understands the magnitude of effort inherent in making such changes. That said, improvements in the Medicare Advantage program, in the various accountable care organization programs, and in other payment or delivery system innovations currently in place can be starting points for this work. In addition, serious attention must be given to new innovations, for example, changing how hospitals are paid and giving providers incentives to manage the cost of medications. The Commission believes that the culmination of the changes we have outlined will provide the Congress and the American people with the opportunity to better predict and manage the long-term cost and quality of the Medicare program.

Although this report sets out a vision for the direction for Medicare payment in the future and makes recommendations for needed changes in today's Medicare payment systems, the Commission realizes that the Congress and CMS are currently coping with the profound challenges facing Medicare and the entire health care system as it contends with the reality of the coronavirus pandemic. The health care system and, most importantly, the individuals caring for the victims of the pandemic need our support and the resources to do their jobs. We will provide whatever advice and assistance that we can at this time to the Congress and CMS as the Medicare program adapts to today's realities. In the future, we will attempt to take lessons learned from today's experience into our assessments of Medicare's payment systems as we help the Congress grapple with the difficult task of controlling the growth of Medicare spending while preserving beneficiaries' access to high-quality care and providing sufficient payment for efficient providers.

Sincerely,

A handwritten signature in dark ink, reading "Francis J. Crosson M.D.", with a stylized flourish at the end.

Francis J. Crosson, M.D.

Enclosure

PFE000198

Acknowledgments

This report was prepared with the assistance of many people. Their support was key as the Commission considered policy issues and worked toward consensus on its recommendations.

Despite a heavy workload, staff members of the Centers for Medicare & Medicaid Services and the Department of Health and Human Services were particularly helpful during preparation of the report. We thank Michelle Cruse, Liz Goldstein, Jennifer Lazio, Blake Pelzer, John Pilotte, Cheri Rice, Abby Ryan, Tiffany Swygert, and Pete Welch.

The Commission also received valuable insights and assistance from others in government, industry, and the research community who generously offered their time and knowledge. They include Michael Bagel, Colin Baker, Greg Berger, Ashley Czin, Richard Fuller, Jane Galvin,

Stuart Hammond, Rick Kronick, Kathy Lester, Chris Lovell, Rob Mechanic, David Muhlestein, Cori Uccello, and Rebecca Yip.

Once again, the programmers and staff at Social & Scientific Systems, a DLH Holdings Corp. Company, provided highly capable assistance to Commission staff. In particular, we appreciate the hard work of Michael Brown, Po-Lun Chou, John Crouch, Daksha Damera, Darya Leyzarovich, Sravani Mallela, Sane Maphungphong, Shelley Mullins, Lorena Ortiz, Cindy Saiontz-Martinez, Allen Selwyn, and Susan Tian.

Finally, the Commission wishes to thank Hannah Fein, Mary Gawlik, and Melissa Lux for their help in editing and producing this report. ■

Table of contents

Acknowledgments	v
Executive summary	xi
Chapters	
1 Realizing the promise of value-based payment in Medicare: An agenda for change	3
Introduction	3
Medicare Advantage and accountable care organizations could provide a foundation for expanding value-based payment	4
The problem of fee-for-service payment	5
A need for change	6
Future vision for the Medicare program	7
2 Challenges in maintaining and increasing savings from accountable care organizations	13
Background	17
Proposed strategies to increase ACO savings	19
Potential for unwarranted shared savings from patient selection	25
Use of NPI for assignment would improve benchmark validity and reduce unintended incentives	37
3 Replacing the Medicare Advantage quality bonus program	47
Quality in Medicare Advantage is difficult to evaluate and the quality bonus program is flawed	51
Design of the new MA–VIP addresses flaws in the current MA quality bonus payment system	55
Illustrative scoring and payment adjustments under the MA–VIP model	70
Replacing the Medicare Advantage quality bonus program with a new value incentive program	86
4 Mandated report: Impact of changes in the 21st Century Cures Act to risk adjustment for Medicare Advantage enrollees	95
Background on Medicare Advantage payments and risk adjustment	99
Changes required by the 21st Century Cures Act	100
Impacts of changes to CMS’s risk adjustment model for Medicare Advantage	101
Summary	114
5 Realigning incentives in Medicare Part D	119
Background	123
Restructuring Part D to restore incentives to manage spending	129
Other modifications to Part D associated with a restructured benefit	141
Recommendations for a restructured Part D benefit	151
6 Separately payable drugs in the hospital outpatient prospective payment system	163
Background	167
Identifying drugs that should be separately payable in the OPPS	169
Considering the criteria used in various Medicare payment systems for the OPPS	172
How long should a drug be separately payable?	174
Summary	175

7	Improving Medicare’s end-stage renal disease prospective payment system	181
	Background	185
	Current payment for new ESRD drugs under the ESRD PPS	186
	Eliminating the TDAPA for new drugs in an existing ESRD functional category	188
	Current payment for low-volume and rural dialysis facilities	190
	Improving the adequacy of payments for low-volume and isolated facilities	195
Appendix A	Commissioners’ voting on recommendations	211
Acronyms		217
More about MedPAC		
Commission members		221
Commissioners’ biographies		223
Commission staff		227

Executive summary

PFE000204

Executive summary

As part of its mandate from the Congress, each June the Commission reports on refinements to Medicare payment systems and issues affecting the Medicare program, including broader changes in health care delivery and the market for health care services. In the seven chapters of this report, we consider:

- ***Realizing the promise of value-based payment in Medicare: An agenda for change.*** The Commission outlines a multiyear effort to lay out a strategic direction for Medicare payment policy and delivery system design that broaden the use of value-based payment.
- ***Challenges in maintaining and increasing savings from accountable care organizations (ACOs).*** The Commission evaluates past savings, examines strategies to increase savings, and recommends a technical change that will reduce the risk that program vulnerabilities might result in unwarranted shared savings payments to ACOs.
- ***Replacing the Medicare Advantage quality bonus program.*** Medicare's quality bonus program (QBP) for assessing and rewarding quality performance in the Medicare Advantage (MA) program is flawed and not consistent with the Commission's principles for quality incentive programs. In the June 2019 report to the Congress, we introduced an alternative MA value incentive program (MA-VIP). In this report, the Commission recommends that the Congress replace the QBP with an MA-VIP that includes five key design elements.
- ***Mandated report: Impact of changes in the 21st Century Cures Act to risk adjustment for Medicare Advantage enrollees.*** The 21st Century Cures Act of 2016 directs the Secretary to make several changes to the CMS hierarchical condition category (CMS-HCC) model, which CMS uses to calculate the enrollee risk scores that adjust MA capitated payments. We assess how each of those changes affects the ability of the CMS-HCC model to predict costs for various Medicare beneficiary populations.
- ***Realigning incentives in Medicare Part D.*** The Commission proposes a package of recommendations to reform Part D and realign plan and manufacturer incentives. The recommendations will limit enrollees' out-of-pocket spending; help restore the role of risk-based, capitated payments; and eliminate features of the current program that distort market incentives. These changes will better align the incentives in Part D with the interests of the Medicare program and its beneficiaries.
- ***Separately payable drugs in the hospital outpatient prospective payment system.*** Medicare payment systems that bundle multiple services into one payment, such as the outpatient prospective payment system (OPPS), create incentives for providers to be judicious about the cost inputs of the services they provide. Paying for items outside the bundle—such as separately payable drugs—should be done only under certain circumstances, such as when a new drug exhibits clinical superiority over an existing drug. In future work, we will determine other criteria for identifying which drugs should be separately payable.
- ***Improving Medicare's end-stage renal disease prospective payment system.*** The Commission recommends (1) eliminating the payment adjustment for certain new drugs and (2) replacing the separate low-volume and rural payment adjustments with a single payment adjustment—a low-volume and isolated payment adjustment—that will protect isolated, low-volume dialysis facilities that are critical to ensure beneficiary access.

Although this report sets out a vision for the direction of Medicare payment systems in the future and makes specific recommendations for needed changes in today's Medicare payment systems, the Commission realizes that the Congress and CMS are currently coping with the profound challenges facing Medicare and the entire health care system as they contend with the reality of the coronavirus pandemic. We will provide whatever advice and assistance that we can at this time to the Congress and to CMS as the Medicare program adapts to today's realities. In the future, we will attempt to take lessons learned from today's experience into our assessments of Medicare's payment systems as we help the Congress grapple with the difficult task of controlling the growth of Medicare spending while preserving beneficiaries' access to high-quality care and providing sufficient payment for efficient providers.

Realizing the promise of value-based payment in Medicare: An agenda for change

In Chapter 1, the Commission outlines a multiyear effort to establish a strategic direction for Medicare payment policy and delivery system design that could be implemented by the Congress and CMS. This work will be aimed at identifying changes that broaden the use of value-based payment (which characterizes methods of paying for health care services that provide stronger incentives than fee-for-service to control overall costs while maintaining or improving quality) by encouraging more providers to organize into “accountable entities.” Such entities would be capable of receiving payments from Medicare and accepting accountability for both the cost and the overall health of a group of beneficiaries. Medicare Advantage and accountable care organizations could serve as vehicles to broaden the use of value-based payment, but both programs need to be improved to realize that potential. This work will be guided by the same fundamental principles that serve as the foundation for all of our policy development: ensuring that beneficiaries have access to high-quality care in an appropriate setting, paying providers equitably and giving them incentives to supply efficient and appropriate care, and assuring the best use of the taxpayer dollars that finance most of Medicare’s spending.

The Commission contends that policymakers will need new approaches to both how Medicare pays providers and how services are organized and delivered to address the currently unsustainable trends in Medicare spending. In 2018, Medicare accounted for 3.6 percent of the country’s gross domestic product, and that figure will grow to 4.7 percent by 2027. As the population ages, the number of workers per Medicare beneficiary is expected to decline—from 3.0 in 2019 to a projected 2.5 in 2029—making the financing of the program more challenging. For example, the program’s Part A trust fund is projected to exhaust its reserves in 2026, which will force Medicare to sharply reduce payment rates for hospitals and other Part A providers unless policymakers take some other action. These trends could result in dramatic changes to the Medicare program and its financing if deliberate changes are not made to how Medicare pays for care and to how care is organized and delivered.

Challenges in maintaining and increasing savings from accountable care organizations

CMS has made it a priority to move more Medicare beneficiaries into alternative payment models in which providers are responsible for the cost and quality of care. One such model is the accountable care organization (ACO). ACOs are now responsible for 23 percent of Medicare beneficiaries with both Part A and Part B coverage. Given the rapid growth in ACOs, it is important to evaluate whether they are generating savings for the Medicare program and thus helping make the program more sustainable. In Chapter 2, the Commission evaluates past savings, examines strategies to increase savings, and recommends a technical change that will reduce the risk that program vulnerabilities might result in unwarranted shared savings payments to ACOs that exceed the rate of savings achieved to this point.

To date, ACOs have generated modest savings, with most evaluations estimating 1 percent to 2 percent reductions in spending from existing ACO models. Some have expressed a concern that the ability of Medicare ACOs to achieve savings has been limited because key constituencies are not sufficiently engaged with ACOs and have incentives that run counter to those of ACOs. CMS and others have expressed an interest in trying to enhance ACOs’ ability to generate savings by creating greater engagement with beneficiaries and specialists, reducing hospital incentives to increase services, and aligning incentives for ACOs and prescription drug use under Part D. However, all of these strategies involve implementation challenges.

Because Medicare savings from Medicare Shared Savings Program ACOs have been relatively small thus far (although still greater than most care coordination demonstrations), there is a risk that those savings could be eroded, or even completely offset, by unwarranted shared savings payments. Patient selection in ACOs could result in unwarranted shared savings payments, whether the selection is intentional or not. For example, if high-cost beneficiaries are disproportionately shifted out of an ACO in its performance year—while remaining in the baseline years—performance-year spending will decrease in relation to the ACO’s benchmark. This phenomenon could occur if clinicians with high-cost beneficiaries bill under a taxpayer identification number (TIN) that is not part of the ACO or if a clinician bills for patients with low spending under the ACO’s TINs and bills for patients with higher spending relative to their risk score under a non-ACO TIN.

The Commission does not believe widespread patient selection occurred in the program's early years. However, the current system allows an ACO to strategically change the composition of its TINs to increase the likelihood of receiving unwarranted shared savings relative to benchmarks, creating a vulnerability for the Medicare program.

To reduce the incentives to select patients and providers, and to reduce the potential mismatch between the clinicians considered in an ACO's baseline years and its performance years, the Commission recommends that the Secretary determine an ACO's historical baseline spending using the same national provider identifiers that are used to compute the ACO's performance-year spending. While there will always be some shared savings payments due to random variation, we should minimize opportunities for unwarranted shared savings payments due to intentional favorable provider and patient selection. Properly matching the clinicians included in an ACO's baseline and performance years will allow a more accurate assessment of an ACO's performance and reduce opportunities for unwarranted shared savings.

Replacing the Medicare Advantage quality bonus program

The Commission maintains that Medicare program payments should take into account the quality of care delivered to beneficiaries, and the Commission has formalized a set of principles for designing Medicare quality incentive programs. Medicare's quality bonus program (QBP) for assessing and rewarding quality performance in the Medicare Advantage (MA) program is not consistent with these principles, and in Chapter 3 we recommend replacing it with a new quality program: the MA value incentive program (MA-VIP).

In our June 2019 report to the Congress, we outlined multiple significant flaws in the QBP program. Those flaws must be addressed so Medicare can have confidence that the MA program encourages and appropriately rewards high quality in a manner that ensures that program dollars are wisely spent. In 2019, MA's QBP cost \$6 billion and is projected by the Congressional Budget Office to cost \$94 billion over 10 years.

The Commission recommends that the Congress replace the QBP with an MA-VIP that includes the following five key design elements:

- ***Scores a small set of population-based measures.***
The measure set would be tied to clinical outcomes as well as patient/enrollee experience.
- ***Evaluates quality at the local market level.***
Evaluating MA plan quality at the local market area level provides information about the quality of care delivered in the localities in which beneficiaries seek and receive care.
- ***Uses a peer-grouping mechanism to account for differences in enrollees' social risk factors.***
Comparing performance among groups of beneficiaries (e.g., fully dual-eligible beneficiaries) with similar characteristics accounts for social risk factors without masking disparities in plan performance, as would be the case if measure results themselves were adjusted by population social-risk characteristics.
- ***Establishes a system for distributing rewards with no "cliff" effects.*** The use of continuous performance-to-points scales allows plans that improve to earn points and avoids the cliff effect, whereby only those plans achieving a certain level of quality receive bonuses.
- ***Distributes plan-financed rewards and penalties at the local market level.*** The MA-VIP redistributes a pool of dollars (made up of a percentage of plan payments within the market areas) as rewards and penalties based on a plan's performance compared with the market area's other plans.

To test the proof of concept of the MA-VIP design, we modeled a prototype MA-VIP using currently available data. In stratifying results by peer groups, the MA-VIP accounts for differences in social risk factors of plan populations and allows plans the potential to earn more rewards for higher quality care provided to populations identified by the presence of certain social risk factors. Our results indicated that an MA-VIP was feasible. An illustrative withhold of 2 percent of payments yielded small penalties and rewards for each peer group for most parent organizations in a market area. To drive quality improvement, policymakers would need to choose an appropriate amount of payment to fund the reward pool and an effective performance-to-points scale methodology.

Mandated report: Impact of changes in the 21st Century Cures Act to risk adjustment for Medicare Advantage enrollees

In Chapter 4, the Commission responds to a mandate in the 21st Century Cures Act that directs it to evaluate the impact of the changes CMS has made to the CMS hierarchical condition category (CMS–HCC) model that is used to risk adjust payments in the MA program.

The Medicare program pays managed care plans that participate in MA a monthly capitated amount to provide Medicare-covered services to its Medicare enrollees. Payment for each enrollee has two parts: a base rate and a risk score. The base rates vary by county, and the base rate for a given county reflects the payment for an MA enrollee in that county with the health status of the national average beneficiary in fee-for-service (FFS) Medicare. The risk score indicates how costly the enrollee would be expected to be in FFS Medicare, relative to the national average FFS beneficiary.

The 21st Century Cures Act of 2016 directs the Secretary to make or consider several changes to the CMS–HCC model, which CMS uses to calculate the risk scores used to adjust MA capitated payments for enrollees. CMS has implemented the changes incrementally: different adjustments for full-benefit and partial-benefit dual-eligible beneficiaries in 2017; adjustments for mental health and substance abuse disorders and chronic kidney disease in 2019; and adjustments for the number of beneficiaries' conditions in 2020.

We have evaluated the impact of the changes that CMS has made to the CMS–HCC model (and the use of two years of diagnosis data, which CMS has not yet implemented) and found the following:

- Each change produces accurate payment adjustments for groups that have characteristics defined by variables in the model.
- Making distinctly different adjustments for full-benefit dual-eligible beneficiaries and partial-benefit dual-eligible beneficiaries eliminates systematic underpayments for the full-benefit dual-eligible beneficiaries and systematic overpayments for the partial-benefit dual-eligible beneficiaries that had occurred in previous models that did not distinguish between these two populations.

- Adding variables to the CMS–HCC model for mental health and substance abuse disorders and chronic kidney disease improves how accurately the model adjusts payments for beneficiaries who have those conditions. However, adding such variables to the CMS–HCC model can provide additional opportunities for MA plans to increase revenue by coding more medical conditions.
- Adding indicators for the number of medical conditions for each beneficiary improves the model's accuracy in adjusting payments for beneficiaries who have no conditions indicated in the model and those who have many conditions.
- Using two years of diagnosis data to determine beneficiaries' conditions is a straightforward and effective method for addressing problems related to differences in coding intensity of medical conditions between MA and FFS Medicare.
- All of the models produce underpayments for beneficiaries with very high levels of Medicare spending and overpayments for those with very low levels of Medicare spending. These payment inaccuracies have been a persistent issue for MA risk adjustment.

We commend the progress that CMS has made in implementing the changes to the CMS–HCC model mandated by the 21st Century Cures Act. We encourage CMS to continue its work on this issue to complete the requirements in the 21st Century Cures Act by the mandated date of January 1, 2022.

Realigning incentives in Medicare Part D

In Chapter 5, the Commission proposes a package of recommendations to reform Part D to limit enrollees' out-of-pocket (OOP) spending; realign plan and manufacturer incentives to help restore the role of risk-based, capitated payments; and eliminate features of the current program that distort market incentives. These reforms will better align the incentives in Part D with the interests of the Medicare program and its beneficiaries. The package of recommendations builds on the major changes the Commission recommended in 2016 to Part D's benefit structure that would have plan sponsors bear more financial risk for their enrollees' drug spending while, at the same time, providing sponsors with greater flexibility to use formulary tools. Changes in law and the expanded use of

high-priced drugs since that time have further eroded the competitive incentives for cost control and have made our new package of recommendations even more crucial.

We recommend restructuring Part D in the following ways:

- For spending below the catastrophic threshold, there would be a standard benefit for all enrollees in which plans would become responsible for 75 percent of spending between the deductible and the catastrophic threshold, with enrollees responsible for the remaining 25 percent through cost sharing. (The proposal would eliminate the manufacturers' coverage-gap discount that currently applies to enrollees without the low-income subsidy (LIS) and remove the coverage gap for LIS enrollees. Because cost sharing for LIS enrollees is limited to nominal copayments, Medicare's LIS would cover most or all of those enrollees' cost sharing.)
- For spending above the catastrophic threshold, the restructured benefit would provide enrollees with greater financial protection by adding an annual cap on beneficiaries' out-of-pocket (OOP) costs. The policy would shift insurance risk from Medicare to plan sponsors and drug manufacturers. Plan sponsors would be liable for more spending in the catastrophic phase than the current 15 percent. A new manufacturers' discount of at least 30 percent would be more likely to apply to drugs and biologics that command high prices, potentially acting as a drag on price growth. (The discount could be structured so that if prices of drugs that were subject to the discount increased faster than a benchmark, the discount rate would increase commensurately.)

The reduction in reinsurance payments and increase in plan liability for spending in the catastrophic phase would be phased in during a transition period so that plan sponsors could adjust to the new distribution of risk. The other elements of the new benefit structure—eliminating the coverage gap, establishing a new discount program in the catastrophic phase, and adding an annual cap on beneficiary OOP costs—would be implemented without a transition.

There are several consequences and actions that would result from these reforms. Sponsors would incorporate lower expected Medicare reinsurance subsidies and higher expected benefit liability into plan bids. Because

Medicare's overall subsidy of basic benefits would remain at 74.5 percent, Medicare's capitated payments to plans would increase to incorporate their new higher benefit liability.

It would be critically important for CMS to recalibrate Part D's risk adjustment model to reflect the increased plan liability. The proposed reforms would result in higher capitated payments for all enrollees, with a larger impact, in dollar terms, for LIS beneficiaries. Given the structure of the risk adjustment model, we believe that CMS would be able to recalibrate the model to ensure that overall payment rates would be adequate for both LIS enrollees and other Part D beneficiaries and for smaller plan sponsors that enroll a higher share of LIS beneficiaries.

Finally, because plans will hold greater insurance risk under the reform, policymakers could consider making the Part D risk corridors more generous to temporarily provide plan sponsors with greater protection during a transition to the new benefit structure. Policymakers could also consider different risk-sharing percentages in the corridors to increase plans' aggregate stop-loss protection. While the enhanced protection would be available to all plans, in practice, the protection would be particularly valuable for smaller plan sponsors that do not have the scale to spread the insurance risk or the capital to reinsure themselves.

Separately payable drugs in the hospital outpatient prospective payment system

In Chapter 6, the Commission specifically considers separately payable drugs in the hospital outpatient prospective payment system (OPPS), although the issues we consider in the chapter have broader implications.

The unit of payment in the OPPS is the primary service (the reason for the visit) coupled with the ancillary items provided with the primary service. That is, the OPPS typically packages the cost of ancillary items into the payment rate of the related primary service. Combining a primary service and related ancillary items into a single payment unit encourages efficiency because the combination of inputs used to treat a patient determines whether the provider experiences a financial gain or loss. However, not all ancillary items are packaged.

A category of ancillary items that has grown in importance in the OPPS is drugs covered under Medicare Part B. The OPPS has two distinct policies for paying some drugs separately from primary services: pass-through drugs and

separately payable non-pass-through (SPNPT) drugs. The pass-through program is intended to provide adequate payment to hospitals for drugs that are relatively costly and new to the market. In contrast, the SPNPT program is intended to provide adequate payment for relatively high-cost drugs that are already established in the drug market. Under both policies, each drug has its own payment rate. Total Medicare spending (combined program spending and beneficiary cost sharing) for pass-through drugs and SPNPT drugs has grown rapidly, increasing from \$5.1 billion in 2011 to \$12.9 billion in 2018. Most of that growth in drug spending—82 percent—was for cancer treatment drugs.

The current criteria for both pass-through drugs and SPNPT drugs have been in place for more than 15 years. We are concerned that the criteria for eligibility under both policies do not strike an appropriate balance between promoting innovation and maintaining pressure on providers to be efficient. Both policies use cost criteria to identify drugs for program eligibility. The cost criteria are different between the programs, but we are concerned that both allow eligibility for drugs that should be packaged. Also, neither policy requires drugs to show that they are clinically superior to competing drugs, even though a requirement for clinical superiority implicitly encourages innovation. As a result, Medicare could pay separately for a drug no more effective than an existing product, even when the cost of the existing product is reflected in the OPPOS payment—resulting in double payments by Medicare.

At this point in our analysis, we conclude that an effective system of separately payable drugs should have two features:

- Some drugs should be paid separately because they are not ancillary. These drugs are the purpose for a visit, are high cost, treat a condition, and are usually administered by infusion.
- Drugs should show clinical superiority over other drugs to have separately payable status. A clinical superiority requirement is vital to prevent double payments by Medicare.

In future work, we will perform analyses to determine other criteria for identifying drugs that should be separately payable. We will also perform analysis to determine the parameters for those criteria.

Improving Medicare's end-stage renal disease prospective payment system

Medicare pays dialysis facilities under a prospective payment system (PPS) that is based on a bundle of services that includes end-stage renal disease (ESRD) drugs (including biologics), clinical laboratory tests, and other items and services. In Chapter 7, the Commission recommends two changes to current payment policy.

First, the Commission recommends that the Congress direct the Secretary to eliminate the transitional drug add-on payment adjustment (TDAPA) for new drugs that are in an existing ESRD functional category already included in the payment bundle. Eliminating the TDAPA would (1) maintain the structure of the ESRD PPS and avoid the introduction of incentives to unbundle services covered under the PPS and (2) create pressure for drug manufacturers to constrain the growth of prices for new and existing ESRD drugs. At market entry, such new drugs would be included in the ESRD PPS bundle without an update to the base payment rate. As new products are added to the bundle and diffused into medical practice, it will be important to monitor the use of ESRD drugs, changes in beneficiaries' outcomes, and the alignment of Medicare payments with providers' costs to evaluate whether a change in the bundled payment is warranted.

Second, the Commission recommends that the Secretary replace the current low-volume payment adjustment (LVPA) and the rural adjustment with a single payment adjustment for dialysis facilities that are isolated and consistently have low volume—where low-volume criteria are empirically derived. The Commission believes that neither the current LVPA nor the current rural adjustment accurately targets facilities that are both critical to beneficiary access and have high costs warranting a payment adjustment.

The Commission modeled a policy—the low-volume and isolated (LVI) adjustment—under which facilities that are low volume and isolated are defined based on both a facility's distance from the nearest facility and total treatment volume. In 2017, the illustrative LVI policy would have applied to 575 freestanding and hospital-based dialysis facilities, compared with 336 facilities receiving the current LVPA and 1,257 facilities receiving the rural adjustment. The LVI policy would not apply to facilities that furnish a high volume of treatments because their economies of scale generally result in lower average treatment costs compared with low-volume facilities.

The LVI policy would also not apply to facilities that are in close proximity to another dialysis facility since such facilities are not the sole providers of dialysis services in their communities and thus are not critical to maintaining

access to care. Overall, the LVI policy would better target payment adjustments to the facilities that are most important for maintaining access to dialysis services and would improve the value of Medicare's spending. ■

C H A P T E R

1

**Realizing the promise
of value-based payment
in Medicare:
An agenda for change**

PFE000214

C H A P T E R

1

Realizing the promise of value-based payment in Medicare: An agenda for change

Introduction

The Commission contends that the growth in spending for the Medicare program poses a significant challenge for the federal government. In 2018, Medicare accounted for 3.6 percent of the country's gross domestic product, and that figure will grow to 4.7 percent by 2027 under current policies. Most of this growth (70 percent) is due to increases in per capita spending (Congressional Budget Office 2019). The expected growth in per capita spending primarily reflects continued growth in payment rates rather than growth in service use. As the population ages, the number of workers per Medicare beneficiary is expected to decline—from 3.0 in 2019 to a projected 2.5 in 2029—making the financing of the program more challenging. The program's Part A trust fund, which pays for services such as inpatient care and post-acute care, is projected to exhaust its reserves in 2026, which will force Medicare to sharply reduce payment rates for Part A providers unless policymakers take some other action (Boards of Trustees 2019). A growing share of program spending—for Part B and Part D benefits—is paid for by general revenues, which are partly financed by deficit spending (Medicare Payment Advisory Commission 2020). Without deliberate changes to the program, this growth in spending could result in dramatic changes to the Medicare program and/or its financing.

The Commission contends that policymakers will need to address this unsustainable trend by changing both how

Medicare pays providers and how services are organized and delivered. A common element for these changes should be the use of value-based payment (VBP), which characterizes methods of paying for health care services that provide stronger incentives to control overall costs than traditional fee-for-service (FFS) while maintaining or improving quality.

This chapter outlines a multiyear Commission effort to establish a strategic direction for Medicare payment policy and delivery system design that could be implemented by the Congress and CMS. This work will be guided by the same fundamental principles that serve as the foundation for all of our policy development: ensuring that beneficiaries have access to high-quality care in an appropriate setting, paying providers equitably and giving them incentives to supply efficient and appropriate care, and ensuring the best use of the taxpayer dollars that finance most of Medicare's spending. This effort will be aimed at identifying changes that broaden the use of VBP by encouraging more providers to organize into entities (which we refer to here generically as “accountable entities”) that are capable of receiving payments from Medicare that require accepting accountability for both the cost and the overall health of a group of beneficiaries. This accountability includes attention to the quality of care, information that beneficiaries can use to compare the care provided by the entities in their area, the systematic provision of preventive services and early detection of

disease, the avoidance of waste, and the delivery of care at the most appropriate and cost-effective site of service.

Medicare Advantage and accountable care organizations could provide a foundation for expanding value-based payment

Although the traditional FFS program has long been Medicare’s primary payment mechanism, about 65 percent of the beneficiaries who have both Part A and Part B coverage are now in two other payment models that have stronger incentives to manage overall spending:

- Almost 24 million beneficiaries (about 42 percent of all beneficiaries with Part A and Part B) are enrolled in Medicare Advantage (MA) plans, which receive capitated payments to provide the Part A and Part B benefit package.
- About 13 million beneficiaries (about 23 percent of all beneficiaries with Part A and Part B) are assigned to accountable care organizations (ACOs), which are groups of FFS providers that have incentives to control overall spending and improve quality.¹

The MA and ACO programs could serve as vehicles to broaden the use of VBP in Medicare, but both programs need to be improved before they can realize that potential. While these programs may be capable of reducing spending relative to the FFS program, whether they actually produce substantial savings depends heavily on how they are structured. For example, 82 percent of MA plans indicate in their bids that they can provide the Part A and Part B benefit package at a lower cost than the FFS program, but the current MA program nevertheless increases overall program spending relative to FFS because MA benchmarks are substantially higher than FFS costs in some areas, many plans receive rebates and quality bonuses, and plans can receive higher payments by submitting more diagnosis codes (Medicare Payment Advisory Commission 2020).

The Commission asserts that broader acceptance in Medicare of accountability for overall costs and outcomes will require improvements to both the existing ACO and MA models. Researchers who have studied the various ACO programs that Medicare has operated over the past 15 years have typically found that they generated modest

savings, on the order of 1 percent to 2 percent of total spending. Whether ACOs will produce larger savings in the future is unclear. CMS has made changes to the largest ACO program—the Medicare Shared Savings Program (MSSP)—that have some positive elements, such as encouraging ACOs to bear more financial risk, but on balance the changes may, in fact, reduce savings for Medicare. In this report, we make a recommendation to protect program savings generated by ACOs in the MSSP by using national provider identifiers instead of tax identification numbers to calculate both performance-year and baseline-year spending.

The Commission plans to conduct further analysis to identify specific policy changes that will improve ACOs and ACO-like models. Any policy changes that we might recommend would be aimed at making ACOs more effective; changes that would, for instance, simply increase funding for ACOs or encourage the creation of ineffective ACOs would provide little, if any, incremental value. Two examples of areas where additional work may be needed (which are discussed in more detail later, in this report's chapter on ACOs) illustrate the complex challenges involved:

- ACOs may be more effective in the longer term if they also have incentives to manage the use of costly prescription drugs. ACOs are currently responsible for the cost of Part A and Part B services only, which includes physician-administered drugs covered under Part B, but does not include outpatient prescription drugs covered under Part D. However, making ACOs more accountable for outpatient prescription drug costs would be challenging because a separate group of entities (Part D plans) already has some financial responsibility for those costs.
- ACOs may be more effective if they have the understanding and support of beneficiaries, who usually do not know that they have been assigned to an ACO and may not be aware of the potential benefits of better-coordinated care. Beneficiaries might be more engaged with ACOs if there were changes to Medigap coverage of out-of-pocket costs and/or financial incentives from ACOs that would encourage beneficiaries to receive care from ACO providers.

Any changes that we might recommend in these and other areas would be intended to increase the chance that these models will be successful. As models improve, we would

support Medicare increasing incentives for providers to participate in them and improve delivery of care.

This work also includes improving the accountability of MA plans to the program and beneficiaries. Over the past several years, we have highlighted numerous shortcomings in the system that Medicare uses to reward plans with high quality ratings. In this report, we recommend the adoption of a new MA value incentive program that would reduce program spending and give beneficiaries better information about the quality of the plans in their area. In the future, we may examine other important aspects of the MA program, such as the benchmarks that help determine plan payment rates and the risk adjustment system.

The Commission may also explore ways to expand the use of value-based payment outside of the scope of the current ACO and MA programs. For example, there is some concern that hospitals have relatively weak incentives, or actually counterincentives, to reduce program spending under the ACO approach. One potential alternative that could give hospitals stronger incentives, but also raises challenging design issues, would be for Medicare to pay hospitals using global budgets that cover all of their inpatient and outpatient services. The state of Maryland is currently testing the use of global budgets for its hospitals in a demonstration under which the hospitals are paid on an FFS basis, but their rates are adjusted to ensure that their overall payments equal a predetermined amount. However, the demonstration's effects have been mixed (hospital spending has decreased, but the effect on overall spending is unclear) and Maryland's approach would be difficult to use at the national level because the state's hospital payments are much higher than traditional Medicare payments. Another issue that may deserve further examination is the possibility of expanding the use of other payment models, such as bundled payments for certain episodes of care, and the need to ensure that those models are well integrated with ACOs.

The problem of fee-for-service payment

Beyond improving the current ACO and MA models, the Commission asserts that, where possible, Medicare should look for ways to further evolve away from the traditional FFS model by identifying policy changes that would dampen its incentives to provide more services while, at the same time, maintaining or improving quality.

Medicare has already made significant efforts to reduce the incentives to provide more services:

- Many FFS payment systems use prospective payments and bundle the payments for related services into a single rate. For example, Medicare pays hospitals a fixed amount for many condition-based episodes of service (through the diagnosis related groups (DRGs) used for inpatient services and the ambulatory payment classifications used for many outpatient services), pays for home health on a per episode basis, and pays for skilled nursing care and most hospice care using daily rates. This approach gives providers an incentive to deliver care efficiently by constraining costs within the episode of service, but it does not limit the number of episodes provided and, to the extent that payments for certain episodes are profitable, could actually spur the provision of unnecessary services.
- Medicare pays private insurers in the MA program a monthly prospective payment for each enrollee. Some plans, in turn, pay delivery system intermediaries (such as an integrated delivery system) a prospective payment for each enrollee. This approach is one example of how providers can be paid using prospective global payment, sometimes referred to as "capitation payment." However, most plans pay providers on a traditional FFS basis. Consideration could be given for Medicare to encourage plans to increase the use of such global payments to providers. One potential benefit of global payments is that providers would have more predictable revenues than they do under FFS, which could mitigate instability during service disruptions such as those that many providers have experienced due to the coronavirus pandemic.
- Medicare pays ACOs based on a variety of payment models, such as bonus-only payments for meeting quality and cost management benchmarks or bonuses based on both upside and downside risk. A small number of ACOs are paid using a capitation model. ACOs may pay individual physician providers based on a variety of payment methods, such as FFS payment, salary with or without volume incentives, or value arrangements such as quality bonuses. Consideration could be given for Medicare to encourage ACOs to pay providers in ways that encourage the delivery of appropriate services

and discourage the provision of unnecessary or inappropriate services.

These and other exceptions to the pure FFS payment model are attempts to constrain the unit cost of services, the number of services provided, or both. These different payment models have had varying levels of success: DRGs have helped constrain Medicare costs, but payments to MA plans have consistently been higher than FFS costs due to the way that Medicare sets plan payment rates, and ACOs have generated only modest savings.

Although the FFS program encourages greater service use, one positive feature of the program is that most of its payment systems use administered prices to pay for services. The use of administered pricing has been helpful in exerting financial pressure on providers and has played a key role in constraining cost growth, especially in recent years as providers have consolidated and negotiated higher commercial rates. For example, Medicare's control over prices is the primary reason its costs have grown more slowly than commercial insurance premiums in recent years. Since Medicare is on a financially unsustainable trajectory, efforts to broaden the use of value-based payment (which focus largely on changing patterns of service use) should be carefully carried out to ensure that they do not inadvertently undermine the program's control over prices.

However, under FFS payment, Medicare beneficiaries may experience significant variability in the quality and appropriateness of services provided and in their resulting outcomes. For example, rates for avoidable hospitalizations and emergency department visits vary across market areas, indicating that there may be opportunities to improve the quality of FFS ambulatory care (Medicare Payment Advisory Commission 2020). There can also be substantial variation in quality within a given type of provider, such as inpatient hospitals (Medicare Payment Advisory Commission 2019). Unfortunately, policymakers now have little ability to compare quality across the FFS, ACO, and MA sectors, and in response, the Commission has supported the use of a small set of outcome, patient experience, and value measures to facilitate those comparisons (Medicare Payment Advisory Commission 2018).

Beneficiaries in the FFS program may also face significant out-of-pocket costs. Traditional Medicare has deductibles for Part A and Part B services, charges copayments or coinsurance for many services, and does not have an

annual cap on beneficiary out-of-pocket costs. As a result, almost 90 percent of beneficiaries have some type of supplemental coverage, such as a Medigap policy, that protects them from high out-of-pocket costs (but also encourages them to use more services). Unlike the FFS program, all MA plans have an annual cap on out-of-pocket costs and cover some Part A and Part B cost sharing. Plans often finance these extra benefits using their MA rebates, which allows many enrollees to obtain some of the same protections as a Medigap policy without having to pay a premium. These extra benefits are one reason that MA plans have become increasingly popular, with many new beneficiaries first enrolling in FFS and then switching to a plan a few years later.

Beneficiaries also experience variable levels of support outside of their direct, physical contact with the delivery system. For example, FFS Medicare does not cover supporting services like transportation, nor does it support the development of preemptive care plans such as population health models that identify gaps in care and seek to close those gaps.

Although administered pricing has helped control spending growth in many parts of the FFS program, it nonetheless has drawbacks. Some services are inevitably mispriced, and payment rates that are too high may encourage inappropriate growth in utilization, as has happened in the past with services such as advanced imaging, therapy in skilled nursing facilities, and durable medical equipment. FFS payment also may contain incentives to overuse new services or lack incentives to provide services that do not have a distinct billing code, such as efforts to address social determinants of health. MA plans and some ACO models may have more opportunity to develop innovative care models in these areas.

A need for change

FFS contains inherent incentives for the delivery system to provide more services and thus receive more payments.² The effects of those incentives are not limited to the FFS program; they also affect how MA plans and ACOs are paid (see next paragraph). Medicare has some counterincentives to avoid the provision of unnecessary or inappropriate services, but they need to be strengthened. The FFS system increases Medicare costs, based on higher than necessary use of services and, in some instances,

the provision of care at higher cost sites of service. The incentive to provide more services also potentially exposes beneficiaries to unnecessary health risks, such as hospital-acquired infections, and to the extra out-of-pocket costs of unnecessary or inappropriate services. Delivery systems that provide care coordination across the continuum of care settings are the exception rather than the norm. There are clearly opportunities for Medicare to provide better value given the large amounts that taxpayers and beneficiaries spend on the program. Finally, the current system does not support sufficient accountability or transparency, such as providing beneficiaries with information that compares the quality of care provided by different models such as FFS, health plans, or ACOs.

The Commission asserts that the development of alternative payment models and care delivery models needs to accelerate.³ There have been numerous efforts by the Congress, CMS (most notably through the Center for Medicare & Medicaid Innovation (CMMI)), and the private sector to address these challenges through MA plans, ACOs, and smaller scale payment and delivery models such as Bundled Payments for Care Improvement (which gives providers incentives to reduce the overall costs for an episode of care) and Comprehensive Primary Care Plus (which makes extra payments to primary care practices that provide more extensive care coordination). Despite these efforts, the development of new payment and care delivery models has had relatively little impact on the average beneficiary and has lagged well behind what is possible and desirable. Policymakers should look for ways to make CMMI more effective so that Medicare can better serve the growing needs of its enrollees.

The Commission contends that unless changes are made to how Medicare pays for services, the cost of the Medicare program will become unsustainable for the country, which could necessitate dramatic changes to the Medicare program and/or its financing. The Commission also contends that the quality of the program will be best served if incentives are aligned between Medicare, the delivery system (through accountable entities), and beneficiaries.

Future vision for the Medicare program

Medicare has used an FFS model to pay for services throughout its history. The FFS program continues to play a central role today, even within the MA and

ACO programs. For example, MA plans bid against benchmarks that equal a percentage of FFS spending, and MA plans are allowed to use FFS rates to pay out-of-network providers (instead of the much higher rates that commercial insurers typically have to pay in those situations). Similarly, the benchmarks that determine whether ACOs qualify for shared savings are tied to FFS spending, and Medicare continues to pay the vast majority of providers affiliated with ACOs on an FFS basis. Medicare's FFS rates are also widely used as a reference point or benchmark by other parts of the health care system.

The FFS model is deeply embedded in our health care system and will probably continue to play an important role after new payment and delivery models are developed. For example, policymakers might use FFS rates to inform the determination of funding amounts for accountable entities, accountable entities might pay for out-of-network or referral services on an FFS basis, and Medicare might continue using the FFS model to pay for care in areas that do not have accountable entities, such as rural areas. Policymakers should thus work to improve the FFS model even as they pursue the development of new payment and delivery models.

Nevertheless, the Commission asserts that the use of FFS payment for Medicare services should be replaced, over time and to the degree feasible, by systems that have incentives to:

- reduce Medicare's financial burden on taxpayers and beneficiaries;
- provide all necessary covered services, including preventive services and early disease detection;
- avoid delivering unnecessary, inappropriate, or low-value services;
- control the costs of providing appropriate and necessary services;
- deliver chronic care services through a care model that features care coordination among providers;
- improve the quality of services and the patient experience of care;
- address and coordinate both the medical and nonmedical needs of beneficiaries; and

- embrace the use of new technologies within payment models that have incentives to reduce program spending or improve quality.

As policymakers develop accountable entities, they may need to consider whether Medicare should support the use of value-based payment by specifying the mechanisms that those entities use to pay individual providers. This approach would represent a departure from current Medicare policy. Medicare has typically stayed out of “downstream” payment arrangements that entities such as MA plans and ACOs use to pay their providers: For example, MA plans have flexibility to negotiate their own payment arrangements with providers, and ACOs have flexibility to determine how shared savings payments are allocated among their participating providers.

Policymakers could find it difficult to develop requirements that account for the range of provider types that deliver Medicare services and the variation in local health care delivery systems. Efforts to promote the use of VBP in the commercial sector have had relatively modest effects to date, and CMS might find that developing and administering requirements in this area would be challenging and prone to unintended consequences. Given these concerns, one approach would be for policymakers to focus on giving accountable entities stronger incentives to control costs and improve quality and then rely on those entities to develop the most effective payment arrangements to meet those goals.

However, as Medicare gains experience with value-based payment, policymakers may be able to develop ways to assess and monitor downstream payment arrangements and determine which methods of value-based payment are more effective. If this happens, Medicare could consider creating incentives that encourage accountable entities to use these models more widely, which could lead to a reduction in the provision of inappropriate and unnecessary services, encourage the delivery of preventive and early disease detection services, facilitate better care coordination among providers, and lower beneficiary out-of-pocket costs, thus justifying the added administrative burden.

Under an improved Medicare program, most beneficiaries would be able to receive their care through a variety of accountable entities that have incentives to both control overall costs and improve quality. Medicare would ideally design incentives that encourage beneficiaries to choose one of these entities to receive their care. Medicare could also strengthen providers’ incentive to participate by reducing FFS payment rates for providers that are not part of an accountable entity. The Commission recognizes that, traditionally, the health care delivery system has been slow to change, and as a result, much of Medicare’s payment apparatus remains connected to legacy payment models. However, the coronavirus pandemic has demonstrated that the system is capable of rapid change when circumstances require it to do so. The Commission asserts that the financing challenges facing the program, its beneficiaries, and the taxpayers who fund it require a similar systemic response to ensure Medicare’s ongoing sustainability. ■

Endnotes

- 1 That figure includes approximately 1 million beneficiaries in Maryland's total cost of care program.
- 2 The FFS incentive to provide more services is reinforced by the widespread use of supplemental coverage to cover some or all of Medicare's out-of-pocket costs. Almost 90 percent of beneficiaries have some type of supplemental coverage. A Commission-sponsored study estimated that spending for elderly beneficiaries with Medigap coverage was 33 percent higher than for those with no supplemental coverage, after controlling for demographics, education, income, and health status (Hogan 2009).
- 3 The steps taken by policymakers and health care providers to address the coronavirus pandemic demonstrate that the delivery system is capable of rapid change. Policymakers and researchers will need to evaluate the effects of recent legislative and regulatory changes on Medicare spending and outcomes to determine which policy changes are worth keeping in place once the pandemic has ended.

References

Boards of Trustees, Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds. 2019. *2019 annual report of the Boards of Trustees of the Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds*. Washington, DC: Boards of Trustees.

Congressional Budget Office. 2019. *The budget and economic outlook: 2019 to 2029*. Washington, DC: CBO.

Hogan, C. 2009. *Exploring the effects of secondary insurance on Medicare spending for the elderly*. A study conducted by staff of Direct Research LLC for MedPAC. Washington, DC: MedPAC.

Medicare Payment Advisory Commission. 2020. *Report to the Congress: Medicare payment policy*. Washington, DC: MedPAC.

Medicare Payment Advisory Commission. 2019. *Report to the Congress: Medicare payment policy*. Washington, DC: MedPAC.

Medicare Payment Advisory Commission. 2018. *Report to the Congress: Medicare payment policy*. Washington, DC: MedPAC.

C H A P T E R

2

**Challenges in maintaining
and increasing savings from
accountable care organizations**

R E C O M M E N D A T I O N

- 2** The Secretary should use the same set of national provider identifiers to compute both performance-year and baseline assignment for accountable care organizations in the Medicare Shared Savings Program.

COMMISSIONER VOTES: YES 16 • NO 0 • NOT VOTING 0 • ABSENT 1

CHAPTER

2

Challenges in maintaining and increasing savings from accountable care organizations

Chapter summary

CMS has made it a priority to move more Medicare beneficiaries into alternative payment models in which providers are responsible for the cost and quality of care. One such model is the accountable care organization (ACO). ACOs are now responsible for 23 percent of Medicare beneficiaries with Part A and Part B coverage. Given the rapid growth in ACOs, it is important to evaluate whether ACOs are generating savings for the Medicare program and thus helping make the program more sustainable. Our work evaluates past savings, examines strategies to increase savings, and discusses how the savings are at risk if program vulnerabilities result in unwarranted shared savings payments to ACOs.

ACOs' savings have been modest

To date, ACOs have generated modest savings, with most evaluations estimating 1 percent to 2 percent reductions in spending from existing ACO models. The Medicare ACO savings stem from small reductions in hospital inpatient, hospital outpatient, and post-acute care use. There have also been savings in at least one commercial ACO according to recent evaluations. The Alternative Quality Contract (AQC) ACO in Massachusetts found material gross savings and modest net savings after accounting for incentive payments to ACOs. AQC savings were primarily due to reduced laboratory testing, imaging, and emergency department visits. Some savings were also generated by using lower priced providers. The larger savings in the commercial ACO

In this chapter

- Background
- Proposed strategies to increase ACO savings
- Potential for unwarranted shared savings from patient selection
- Use of NPI for assignment would improve benchmark validity and reduce unintended incentives

should be expected given that the AQC model evaluated is housed in an HMO that—unlike Medicare ACOs—can use prior authorizations to restrict service use and has the ability to steer patients to lower priced providers.

Some have expressed a concern that the ability of Medicare ACOs to achieve savings has been limited because key constituencies are not sufficiently engaged with ACOs and have incentives that run counter to those of ACOs. CMS and others have expressed an interest in trying to enhance ACOs' ability to generate savings by creating greater engagement with beneficiaries and with specialists, reducing hospital incentives to increase services, and creating incentives for ACOs to control prescription drug use under Part D. However, all four of these strategies involve implementation challenges.

Technical change to reduce unwarranted shared savings from patient selection

Because Medicare savings from Medicare Shared Savings Program ACOs have been relatively small thus far (although still greater than most care coordination demonstrations), there is a risk that those savings could be eroded, or even completely offset, by unwarranted shared savings payments. Unwarranted payments can result if there is patient selection in ACOs, whether intentional or not. For example, if high-cost beneficiaries are disproportionately shifted out of an ACO in its performance year while remaining in the baseline years, performance-year spending will decrease in relation to the ACO's benchmark. This selection can occur if high-cost clinicians are removed from the ACO or if clinicians with high-cost beneficiaries bill under a taxpayer identification number (TIN) that is not part of the ACO. A second means of patient selection involves removing just a portion of a high-cost provider's patients from the ACO. The clinician could bill for patients with low spending under the ACO's TINs and bill for patients with higher spending relative to their risk score under a non-ACO TIN.

The Commission does not believe widespread patient selection occurred in the program's early years. However, the current system allows an ACO to strategically change the composition of its TINs to increase the likelihood of receiving unwarranted shared savings relative to benchmarks, creating a vulnerability for the Medicare program.

To reduce the incentives to select patients and providers, and to reduce the potential mismatch between the clinicians considered in an ACO's baseline years and its performance years, the Commission recommends that the Secretary determine an ACO's historical baseline spending using the same national provider identifiers (NPIs) that are used to compute the ACO's performance-year spending. Properly

matching the clinicians included in an ACO's baseline and performance years will allow a more accurate assessment of an ACO's performance and reduce opportunities for unwarranted shared savings. While there will always be some shared savings payments due to random variation, we should minimize opportunities for unwarranted shared savings payments due to favorable provider or patient selection. In other words, ACOs should be rewarded for achieving real savings due to improving patient outcomes and appropriately managing utilization—not for apparent gains that result from unnecessary mismatches between the clinicians included in performance-year and baseline-year (benchmark) calculations. ■

Background

Organizations of providers that agree to be held accountable for the cost and quality of care are called accountable care organizations (ACOs). The goal of ACOs is to create an incentive for providers to control spending growth and improve quality for a population of Medicare fee-for-service (FFS) beneficiaries. Because ACOs are provided with claims data for their beneficiaries, they can theoretically improve care coordination and encourage their beneficiaries to use more efficient providers—though beneficiaries still have the freedom to choose to receive their care from any Medicare-participating provider. Compared with Medicare Advantage (MA) plans, ACOs have fewer tools to control use (e.g., they cannot limit provider networks, cannot require prior authorization), but they also have lower marketing and administrative costs.

Almost a quarter (23 percent) of Medicare beneficiaries with both Part A and Part B coverage are assigned to ACOs. CMS assigns beneficiaries to an ACO if they have a plurality of primary care visits with clinicians who participate in the ACO. Most of these beneficiaries are assigned to ACOs in the Medicare Shared Savings Program (MSSP), a permanent ACO model established through the Affordable Care Act of 2010 (ACA). Since its inception in April 2012, the MSSP has grown rapidly. In 2020, there are 517 MSSP ACOs responsible for the cost and quality of care provided to 11.2 million FFS beneficiaries. Although this chapter focuses on the MSSP, CMS has also operated a series of ACO-related demonstration programs through the Center for Medicare & Medicaid Innovation (CMMI), including separate programs in Maryland and Vermont. In addition, private insurers (including parent organizations of MA plans and commercial insurers) also operate ACOs.

For each ACO, CMS sets a spending target for a beneficiary population assigned to that ACO. This target is called a benchmark. If Medicare spending for care provided to an ACO's assigned beneficiaries is below this benchmark, the ACO can receive "shared savings" payments, which can range from 50 percent to 100 percent of shared savings in different ACO models. If Medicare spending is above the benchmark, the ACO may share liability, depending on its risk arrangement with Medicare. Under a one-sided risk arrangement, the ACO bears no liability for spending exceeding its benchmark. Under a two-sided risk arrangement, the ACO may be liable for

some share of the difference between actual spending and the benchmark. CMS must strike a balance when setting ACO benchmark rules. If CMS sets benchmarks too low, providers could doubt their ability to generate savings and could therefore avoid participating in the program (especially in two-sided risk arrangements). In contrast, if CMS sets benchmarks too high, providers would be able to keep spending under the benchmarks without appreciably altering the provision of care, thereby receiving unwarranted "shared savings" payments. In this scenario, the ACO program would cause overall Medicare spending to increase rather than decrease.

To date, ACOs have generated relatively small savings, but those savings are nevertheless greater than those achieved in most care coordination models in Medicare. We define Medicare savings from an ACO program as savings evaluated against a counterfactual—that is, what spending would have been if the ACO program did not exist. Performance-year savings can be reduced by "shared savings" payments made to the MSSP's participating ACOs to calculate net savings to Medicare. In contrast, CMS's shared savings payments are evaluated relative to the ACOs' benchmarks, not to a counterfactual. Hence, unwarranted shared savings payments can be made if they result from a mismatch between benchmarks and actual spending. Accordingly, ACO models must be designed to minimize opportunities for ACOs to receive unwarranted shared savings payments.

The ACO program has grown rapidly

The MSSP started in 2012 with 114 ACOs in the initial cohort and grew to 561 ACOs by January 2018. In 2019, CMS introduced new MSSP rules, referred to as "Pathways to Success." As of July 2019, there were 518 ACOs in the MSSP (Table 2-1, p. 18), making 2019 the first year in which the number of ACOs leaving the program exceeded the number joining the program.¹ By January 2020, there were 517 ACOs in the MSSP. Despite the decline in numbers of participating ACOs since 2018, the number of assigned beneficiaries in the MSSP has continued to increase every year, with 10.9 million beneficiaries in the program in 2019 and 11.2 million as of January 2020. From 2013 to 2020, the average size of an ACO increased from 14,500 beneficiaries to 21,600 (data not shown).

The Pathways to Success introduced in 2019 created new MSSP models designed to move MSSP ACOs more rapidly to two-sided risk. (See the Commission's *Payment*

**TABLE
2-1****The number of MSSP ACOs increased through 2018 and has since decreased**

	2013	2014	2015	2016	2017	2018	2019*	2020
Beginning of year	220	338	404	433	480	561	518	517
New to program	106 **	123	89	100	99	124	66	53
Left program (previous year)	0	5	23	71	52	43	109	54
Beneficiaries (in millions)	3.2	4.9	7.3	7.7	9.0	10.5	10.9	11.2

Note: MSSP (Medicare Shared Savings Program), ACO (accountable care organization).

*Data as of July 1, 2019. Because of the ACO rule change, in 2019, new ACOs joined in July, not January. Sixty-six ACOs joined in July 2019 and 109 ACOs left the program in the previous year or in 2019 before July 1.

**114 ACOs joined in 2012, the first year of the program.

Source: MedPAC analysis of CMS data.

Basics for more detail: http://medpac.gov/docs/default-source/payment-basics/medpac_payment_basics_19_aco_final_sec.pdf?sfvrsn=0.) Nonetheless, in 2020, most MSSP ACOs remain in one-sided models.

ACO models' savings to date have been modest

Evaluation of various Medicare ACO models and one rigorously evaluated commercial model have shown small ACO savings. Gross savings were larger in the largest commercial ACO program that has undergone a thorough evaluation, but net savings (after incentive payments) were still small. These evaluations all define savings as the difference between actual spending and what spending would have been in the absence of the ACO program (this counterfactual approach is not equivalent to the CMS computation of "shared savings" relative to a benchmark).

Medicare program savings from all ACO models generally have ranged from 1 percent to 2 percent

Over the past 15 years, all of the ACO models evaluated by CMS have generated similar savings, despite key differences in assigning beneficiaries, setting benchmarks, determining comparison groups, and adjusting for risk. Even with these differences, the four early ACO models (the Physician Group Practice Demonstration, the Pioneer ACO demonstration, the initial MSSP model, and the Next Generation ACO model) all appear to have generated modest savings for the Medicare program in the range of 1 percent to 2 percent. (See text box for a history of the savings and incentives of the various ACO models,

pp. 20–21.) For example, our estimate of MSSP savings from 2012 to 2016 showed a 1 percent or 2 percent slower rate of growth for spending on beneficiary populations in MSSP ACOs in 2013 (not accounting for shared savings payments) (Medicare Payment Advisory Commission 2019). Although the estimated savings from these models are modest, they surpass those achieved by a wide variety of care coordination models Medicare has tried. Thus, it is important that these opportunities for program savings be preserved in future ACO models.

However, the latest MSSP model, which began in 2019, is designed to be on balance more favorable to certain ACOs and likely will result in larger "shared savings" payments to participating ACOs given any level of performance. If so, the new MSSP model may not generate any net savings for Medicare, unless the new model has a materially larger effect on service use than did previous ACO models. One concern is that the rules for the new MSSP model create incentives for ACOs to direct resources toward increased diagnostic coding (because risk score increases are now allowed to increase benchmarks) and toward seeking a favorable selection of clinicians and patients (which is easier given regional benchmarks) rather than improving care and reducing unnecessary use of services.

Commercial ACO programs have mechanisms for generating savings that may not be available to the Medicare program

ACOs have become more common within commercial insurance payment models. According to Leavitt Partners, there were 876 commercial ACO contracts in 2019, and

the number has been growing (Muhlestein et al. 2019). Although there are many commercial ACO programs in operation, the most extensively studied commercial ACO program has been the Alternative Quality Contract (AQC) between Blue Cross Blue Shield of Massachusetts and ACOs in the Boston area. It has often been cited as a successful example of how ACOs can operate in the commercial sector. The text box on savings from commercial ACOs (pp. 22–23) summarizes the most recent evaluation of the AQC. Although the evaluation found that the AQC resulted in savings even after incentive payments to the ACOs, that level of savings may not be achieved by Medicare ACOs because Medicare ACOs have fewer tools. For example, savings from switching patients from high-priced to low-priced hospitals would be limited in Medicare because Medicare sets relatively uniform rates for all providers of the same type.

Proposed strategies to increase ACO savings

Some stakeholders have expressed a concern that the ability of ACOs to achieve savings has been constrained because key constituencies have not sufficiently engaged with ACOs. CMS and others have expressed an interest in trying to enhance ACOs' ability to generate savings by creating greater engagement with beneficiaries and specialists, reducing hospital incentives to increase services, and aligning incentives for ACOs and prescription drug plans under Part D. Recent changes in Medicare policy are intended to allow two of these strategies—beneficiary engagement and aligning hospital incentives—to be tested.

Increasing beneficiaries' incentives to engage with an ACO

Initially, ACOs had few tools with which to encourage beneficiaries to become engaged with an ACO. (Beneficiaries are often not aware they are in an ACO and could have difficulty understanding the ACO concept. Engagement with an ACO, therefore, usually translates to engagement with their primary care physician's practice.) Historically, ACOs' primary tool was providing high-quality care and thus convincing beneficiaries that they should continue to see the ACO's primary care physicians. However, beneficiaries often change the physicians they see as their health care needs change or they have issues with their current providers, and about 25 percent of ACO beneficiaries were switched out of their ACO in 2017.

Since 2019, the new MSSP and the proposed CMMI Direct Contracting model have created new tools for beneficiary engagement. ACOs can encourage beneficiaries to consistently use the ACO's primary care practice by providing supplementary benefits such as:

- cash payments of up to \$20 for seeing ACO physicians if the beneficiary is in a two-sided ACO model
- paying for transportation services
- vouchers for chronic disease management programs, wellness programs, or meal programs
- items to support management of chronic disease, such as air-filtering systems or air conditioners
- waiving cost sharing (allowed in the CMMI Direct Contracting model)

ACOs can also have beneficiaries name their primary care physician, which will govern enrollment as long as they have recently used that physician. In a recent proposed rule, CMS also discussed allowing beneficiaries to directly enroll in an ACO similar to beneficiary enrollment in an MA plan (Centers for Medicare & Medicaid Services 2018b). However, some commenting on the rule suggested that the ACO concept may be difficult to explain to beneficiaries and could create confusion between ACOs and MA plans (Centers for Medicare & Medicaid Services 2018a).

Given the wide range of tools ACOs can now use to engage beneficiaries, the question is no longer whether ACOs have the tools to engage a beneficiary. The question is whether the ACOs believe the cost of the extra benefits (borne by the ACO) will be offset by savings from reduced service use if the patient continues to use ACO clinicians.

Increasing hospitals' incentive to reduce unnecessary service use

On average, hospital-led ACOs have not generated savings in the MSSP (McWilliams et al. 2018, Medicare Payment Advisory Commission 2019). Some have attributed this result to hospitals' lack of incentive to reduce volume. Hospitals may prefer increasing FFS revenue through increasing volume over the opportunity to achieve shared savings through reduced volume and revenue. In addition to insufficient hospital incentives, hospital-led ACOs may generate less savings because their typically large physician staffs each have a small individual incentive to act efficiently since the savings from their personal efforts

History of Medicare accountable care organizations

2005 to 2010: The Physician Group Practice Demonstration

- **Population:** 220,000 beneficiaries at 10 organizations selected by the Secretary
- **Key design features:**
 - Benchmark based on historical spending; benchmark growth based on local competitors' spending growth
 - Hierarchical condition category (HCC) coding growth increased benchmarks
 - One-sided risk (bonus only)
 - Retrospective assignment
- **Ways to obtain "shared savings":**
 - Lower spending growth
 - Increase coding
 - Have local competitors with high spending growth
- **Program savings:** Estimated at 1 percent to 2 percent savings in an average year with net savings (after shared savings payments) of less than 1 percent (RTI International 2012)

2012 to 2016: Pioneer ACO (Center for Medicare & Medicaid Innovation (CMMI) demonstration)

- **Population:** Up to 700,000 beneficiaries in 32 organizations selected by the Secretary (most Pioneer accountable care organizations (ACOs) withdrew from the program before it ended)
- **Key design features:**
 - Benchmark based on historical spending; benchmark growth based on national spending growth rates; evolved to adjust for changes in local prices
 - HCC growth did not affect benchmarks

- One-sided risk (first year) evolving to two-sided risk (bonus and penalty)
- Waiver of three-day skilled nursing facility stay rule
- Beneficiaries could voluntarily align with an ACO
- Prospective assignment
- **Ways to generate "shared savings":**
 - Lower spending growth
 - Opportunities for patient selection were lower in the Pioneer program than in the Medicare Shared Savings Program (MSSP) due to prospective assignment
- **Shared savings:** Initial year savings estimated between 1 percent and 2 percent before shared savings payments and less than 1 percent after shared savings payments (McWilliams et al. 2015)

2012 to 2019: Initial MSSP shared savings model (the MSSP is permanent)

- **Population:** 10.5 million beneficiaries in 561 ACOs by 2018
- **Key design features:**
 - Benchmark based on historical spending, adjusted for national growth in spending and for changes in local prices
 - HCC growth did not increase benchmarks; HCC declines reduced benchmarks
 - Primarily one-sided risk (bonus only)
 - Retrospective assignment
- **Ways to generate "shared savings":**
 - Lower spending growth
 - Use wellness visits to maintain assignment of beneficiaries with low utilization

(continued next page)

History of Medicare accountable care organizations (cont.)

- Random variation can benefit ACOs in one-sided (bonus-only) models
- **Shared savings:** Savings estimates depend on year and methods, but still are generally in the 1 percent to 2 percent range before shared savings payments; near 1 percent after shared savings payments (McWilliams et al. 2018, Medicare Payment Advisory Commission 2019)

2015 to 2019: Next Generation (NextGen) ACO model (CMMI demonstration)

- **Population:** 500,000 beneficiaries in 18 ACOs in 2016
- **Key design features:**
 - Benchmark is primarily based on historical spending, adjusted for national spending growth and local price changes
 - HCC growth can increase benchmarks by up to 3 percent, but a common coding adjustment across ACOs reduces some of the coding growth for NextGen ACOs
 - Two-sided risk (bonus and penalty)
 - Prospective assignment
- **Ways to generate “shared savings”:**
 - Lower spending growth
 - Increase coding faster than the coding adjustment applied to all ACOs
- **Shared savings:**
 - First year evaluation: 1 percent to 2 percent reduction—relative to fee-for-service (FFS) Medicare—before shared savings and approximately 1 percent after shared savings payments (NORC at the University of Chicago 2018)
 - Second year evaluation: The evaluation compared the NextGen model against all other FFS Medicare (including MSSP ACOs) and

found no net savings, perhaps in part due to MSSP savings (NORC at the University of Chicago 2020)

2019 onward: New MSSP model (MSSP is a permanent program)

- **Population:** Total MSSP population 10.9 million beneficiaries in 518 ACOs in mid-2019
- **Key design features:**
 - Benchmarks are a blend of historical and regional spending, and benchmark growth is a blend of national and regional growth
 - Asymmetric risk and rewards favor ACOs
 - Allows HCC coding to increase benchmarks up to 3 percent; unlike Medicare Advantage and NextGen, there will be no across-the-board coding adjustment
 - Annual choice of retrospective or prospective assignment
- **Ways to generate “shared savings”:**
 - Lower spending growth
 - Begin with spending levels lower than others in the market
 - Improve patient mix by changing choice of prospective or retrospective assignment from one year to the next
 - More complete coding
 - Random variation rewards are larger than penalties; therefore, expected shared savings due to random variation is positive, but providers must take risk or have a partner take risk
 - Use wellness visits to maintain assignment of beneficiaries with low utilization
 - Adjust which national provider identifiers bill to ACO taxpayer identification numbers to improve patient selection ■

Savings from commercial ACOs may be difficult to replicate in Medicare ACOs

Blue Cross Blue Shield (BCBS) of Massachusetts instituted a two-sided population-based global budget (or accountable care organization (ACO)) contract, called the Alternative Quality Contract (AQC), for some of its commercial enrollees. The AQC was launched in 2009 with provider organizations that collectively cared for about 20 percent of BCBS's HMO members; by 2013, 85 percent of HMO members and providers in the BCBS network had entered the AQC. HMO enrollees select a primary care physician (who controls referrals to specialists); HMO enrollees are then assigned to that primary care physician's ACO. By 2016, the program had experienced lower growth in spending on medical claims for HMO enrollees relative to a comparison group of HMO enrollees across eight northeastern states. By the eighth year of the contract, growth in medical spending for AQC members relative to the comparison group was reduced by an average of 11.7 percent for enrollees in organizations that entered in 2009, 11.9 percent for those entering in 2010, 6.9 percent for those entering in 2011, and 2.3 percent for those entering in 2012 (Song et al. 2019). These

savings are computed before incentive payments to providers, which were larger in the initial years of the program than in the later years. Therefore, net savings were modest. On net, however, Song and colleagues estimated that, using unadjusted averages weighted by enrollment, reductions in medical claims relative to the comparison group were about 3 percent larger than incentive payments across the different ACO cohorts (Song 2020).

Following are the key findings from the AQC evaluation:

- The AQC was not associated with a reduction in inpatient services.
- The AQC was associated with a reduction in "laboratory testing, certain imaging tests, and emergency department visits."
- The AQC was associated with patients using lower priced sites of care, with approximately 29 percent

(continued next page)

will be shared among the whole organization. Finally, even if a hospital has an incentive to reduce volume, the hospital administrators may be reluctant to make the difficult decisions to reduce the size of their organization. In general, reducing an organization's growth is, by nature, counterintuitive and may not be rewarded by the hospital's board.

While historically, ACOs in one-sided models received only 50 percent of shared savings, CMS has moved toward giving hospitals in ACOs larger incentives to reduce hospital volume. Under two-sided models, shared savings rates rose to 75 percent in the enhanced MSSP model, 80 percent in the Next Generation (NextGen) model, and are proposed to go up to 100 percent shared savings in the Direct Contracting model. To the extent that the problem is a lack of institutional incentive, 100 percent shared savings could help solve the problem.

In addition to institutional incentive issues, the hospital's culture may still be influenced by payments for non-ACO Medicare beneficiaries and commercial patients for whom the hospital receives FFS payments. One notable exception is Maryland, where hospitals have had an all-payer global budget since 2014. This payment model reduces the issue of mixed incentives. However, a recent analysis of Maryland's hospital global budget model suggests that although inpatient use was reduced, it was not clear that net Medicare spending was reduced (Haber et al. 2018, Roberts et al. 2018). In addition, Maryland is unique in that the level of Medicare payments under the global budget is far higher than what Medicare payments would have been under traditional FFS rates. Expanding the Maryland model to other states would be difficult, and it is not clear that overall spending would decline, given the high level of spending in Maryland and the lack of clear findings on changes in Maryland's overall Medicare spending.

Savings from commercial ACOs may be difficult to replicate in Medicare ACOs (cont.)

of savings resulting from using lower priced services rather than fewer services.

- For most ACO cohorts, the savings from reduced service use began to exceed the incentive payments provided to the ACO in the later years.
- Quality of care process measures improved as did outcome measures for hypertension and diabetics' control of glycated hemoglobin.
- A study of differences in AQC performance among lower and higher socioeconomic status groups found: "Quality improved for all enrollees in the Alternative Quality Contract after their provider organizations entered the contract. Process measures improved 1.2 percentage points per year more among enrollees in areas with lower socioeconomic status than among those in areas with higher socioeconomic status. Outcome measure improvement was no different between the subgroups; neither were changes in spending" (Song et al. 2017).

While the results from the AQC are promising, Song and colleagues warned that they may not be generalizable for other ACO arrangements such as Medicare because 29 percent of the AQC savings in early years resulted from using lower cost providers (a price effect) rather than using fewer services (a quantity effect). For example, savings could occur when volume shifts from a higher priced hospital to a lower priced hospital. Similar savings may be more difficult to achieve in Medicare in part because Medicare sets prices administratively. In addition, some have noted that the model is more easily implemented with an HMO population than in more open coverage arrangements, such as preferred provider organizations (PPOs) or Medicare ACOs. In 2016, the AQC expanded to include Massachusetts's BCBS PPO members, and providers have continued to accept two-sided risk for both HMO and PPO members under these contracts. A formal evaluation of results in the PPO context, similar to the evaluation conducted on the HMO model, is under way. ■

Increasing specialist engagement with ACOs

Some stakeholders contend that ACOs need to meaningfully engage specialists in efforts to practice conservatively. Several factors can influence specialists' participation in ACOs, such as the potential to increase their referrals from the ACO's primary care physicians, to share in savings if the ACO reduces spending below its benchmark, and to receive a 5 percent incentive payment from Medicare if the ACO qualifies as an advanced alternative payment model (A-APM) (clinicians with substantial participation in an A-APM receive a payment worth 5 percent of their professional services payments in a lump sum from 2019 through 2024).²

ACOs might want to include specialists as participating physicians because, through incentives, they can influence specialists to practice conservatively and avoid unnecessary services. However, ACOs may not see a need to include specialists because beneficiaries are

mainly assigned to ACOs based on their primary care visits with primary care clinicians. Even if specialists do not participate in an ACO, the ACO can still influence specialists' practice patterns if the ACO's primary care physicians influence referrals to specialists.

Interviews with ACO leaders and focus groups with physicians provide insights into whether ACOs seek to include specialists and how these organizations manage the use of specialty services. These findings come from two sources: (1) interviews that Commission staff conducted in 2018 with leaders of 17 ACOs in 3 states that were participating in the MSSP and NextGen programs and (2) focus groups conducted by Commission staff in 2019 with physicians in markets that have Medicare ACOs.³

Among the ACOs we interviewed, the share of participating specialists varied widely. ACOs led by primary care physician groups may be more selective about their participating physicians than other ACOs and

may not include any specialists, but many of the health system-affiliated ACOs (and one led by a multispecialty group practice) include more specialists than primary care physicians. Health system-affiliated ACOs tend to include all their employed physicians in their organization, which might explain why these ACOs have more specialists than primary care physicians. ACOs that include specialists told us that participating specialists may be less aware than primary care physicians that they are part of an ACO. According to the physician focus groups we conducted, specialists who participate in an ACO may or may not receive a bonus when their ACO produces shared savings. Some specialists felt frustrated that they were not financially rewarded when their ACO reduced spending.

The ACOs interviewed said they use various approaches to manage referrals to specialists. One technique is to encourage primary care clinicians to refer patients to lower cost specialists. For example, one ACO gives its primary care physicians data on how specialists are ranked based on their use of services. According to ACOs, when specialists know that information on their cost and use of services will be shared with primary care clinicians, it gives specialists a strong incentive to change their behavior. Another tool to reduce the cost of specialty care is to give specialists information about their service use (e.g., the number of tests, procedures, and follow-up visits).

Our analysis of physician participation in ACOs found that the share of specialists participating in MSSP and NextGen ACOs in 2018 was similar to the share of specialists among all physicians participating in Medicare. Of physicians participating in MSSP ACOs and NextGen ACOs, specialists accounted for 63 percent and 60 percent, respectively. By comparison, in 2018, 64 percent of all physicians participating in Medicare were specialists. The share of specialists is generally higher in hospital-affiliated ACOs than physician-led ACOs. Among MSSP ACOs in 2018, 65 percent of physicians in hospital-affiliated ACOs were specialists, compared with 50 percent of physicians in physician-led ACOs. The gap is larger among NextGen ACOs: In 2018, in hospital-affiliated ACOs, 63 percent of physicians were specialists, compared with 36 percent in physician-led ACOs. One explanation for the higher share of specialists in hospital-affiliated ACOs could be that these types of ACOs tend to include all of a hospital's employed physicians.

To explore whether MSSP ACOs that earn shared savings share the savings with specialists, we looked at public

websites for a sample of 200 MSSP ACOs from the 2018 performance year. Of those ACOs, 69 (35 percent) had easily searchable websites that listed how they distributed shared savings. On average, those ACOs distributed 58 percent of their savings to providers, although the share distributed to providers varied widely. The remaining ACOs' shared savings went to administration and infrastructure. Only eight ACOs reported how they distributed shared savings among provider types. Six ACOs distributed 60 percent of their shared savings to providers, all of which went to primary care clinicians. One ACO reported that, of the 75 percent of shared savings distributed to providers, 60 percent went to physicians (whether the physicians were primary care physicians or specialists was not specified) and 40 percent went to hospitals. Another ACO reported that it distributed 70 percent of its shared savings to providers; 60 percent went to primary care physicians, 35 percent to specialists, and 5 percent to hospitals.

Although few studies examine the impact of specialists' participation in ACOs on volume and spending, one study found that MSSP ACOs with a high share of primary care physicians were more likely to reduce the number of visits with specialists than ACOs with a high share of specialists (Barnett and McWilliams 2018). These results are consistent with the authors' hypothesis that ACOs with more primary care physicians have a stronger incentive than other ACOs to reduce the use of specialty care because they do not lose FFS revenue when they provide less specialty care. Another study found that independent primary care group ACOs in the MSSP reduced total Medicare spending but independent multispecialty group ACOs did not (McWilliams et al. 2016a, McWilliams et al. 2016b).

Challenges in bringing Part D drug spending into ACO benchmarks

Medicare ACOs are held accountable only for the cost of Part A and Part B services. Notably absent are the costs of outpatient prescription drug spending, even though ACO clinicians directly prescribe medications for their patients.

Despite the important role pharmaceuticals play in treating many conditions, Part D, Medicare's program for outpatient drugs, operates separately from Part A and Part B. Not all beneficiaries in FFS Medicare enroll in Part D, but those who do are enrolled in one of the typically dozens of privately run stand-alone prescription drug plans (PDPs) that operate in their geographic region, and they can change their enrollment decision annually.

Plan sponsors that operate PDPs usually have no direct relationship with prescribers. PDPs must provide access to a broad set of drugs most commonly needed by enrollees as recognized in national treatment guidelines, but the specific medicines included on each plan's formulary or drugs that are assigned preferred cost sharing vary. Part D's payment and enrollment systems are distinct from those of FFS Medicare, and although PDP sponsors bear financial risk for prescription drug spending, they are not at risk for medical spending.

Unlike Medicare ACOs, formal integration of medical and drug spending is common among ACOs with commercial contracts. According to one national survey of ACO executives conducted between 2012 and 2014, 76 percent of ACOs that had at least one commercial contract were held responsible for drug costs in their largest contract (Colla et al. 2015).

Approaches toward integrating medical and drug services

Increased alignment of ACOs and Part D has the potential to create a more comprehensive approach to improving the efficiency of care delivery. However, carrying out such integration would be complex. For example, CMS could include Part D spending in ACO benchmarks without formal collaboration between ACOs and PDPs. Alternatively, CMS could encourage Part D plans to contract with ACOs to reduce drug spending. Both approaches are challenging.

Approach 1: Add Part D spending to the ACO benchmark

Under the first approach, CMS would use past Part D claims for each ACO assignee to project a drug spending benchmark to add to the ACO's Part A and Part B benchmark. ACOs would have the opportunity to share savings if actual spending for combined medical and drug benefits were lower than the projected benchmark. The approach has advantages, notably giving ACOs stronger incentives to evaluate prescription use and spending in their decision-making. However, not every FFS beneficiary chooses to enroll in Part D, so CMS would not have drug claims for all ACO assignees to add to benchmarks. Because Medicare already holds PDPs accountable for some Part D spending through capitated payments, this approach of adding drug spending to the ACO benchmark would separately compensate two sets of providers (PDPs and ACOs) for bearing the same risk. In addition, projecting Part D benchmarks would be difficult. The agency would need to develop methods to attribute

rebates and discounts to individual beneficiaries to reflect their historical net drug spending, and then project forward expected future rebates. A further problem is that this model would not integrate ACO and PDP providers' decision-making regarding formularies and benefit design.

Approach 2: Encourage ACOs to contract with Part D plans

Under a second approach, CMS would encourage and support private collaboration between ACOs and PDPs. In recent years, Medicare ACOs have built partnerships with a number of entities related to prescription drug spending, including PDP sponsors such as CVS Caremark and pharmacy chains such as Walgreens. While they have had mixed success, the general goals of these collaborations include filling gaps in care (e.g., administering flu shots), sharing data, and helping to set targets for and monitor prescription drug adherence. In 2014, SilverScript, CVS Caremark's (now CVS Health) brand of stand-alone Part D plans, announced that it was entering into a shared savings arrangement with several ACOs to lower Part D drug spending for its enrollees (Avalere Health 2014).⁴ The arrangement provided ACO partners with financial incentives to reduce drug spending through one-sided shared savings for Part D costs. According to the announcement, SilverScript would benefit only from lower drug spending, not from lower FFS spending, even if those savings were a consequence of improved medication adherence. SilverScript's collaborations with ACOs appear to have continued at least through 2017 (Brennan 2017). CVS Caremark continues to promote its potential role in improving health outcomes and lowering costs by leveraging its data and the ability to screen for evidence of nonadherence or safety concerns. CVS Caremark's enthusiasm for ACO collaborations suggests that SilverScript reaped some benefits through these partnerships. However, there are currently no published studies on how effective SilverScript's ACO collaboration has been. To the extent that this model works, there may be little for CMS to do other than facilitate the exchange of information.

Potential for unwarranted shared savings from patient selection

Because Medicare savings from MSSP ACOs have been modest thus far (although still greater than most care coordination demonstrations), those savings need to be protected from unwarranted shared savings payments to

Beneficiary assignment in the MSSP

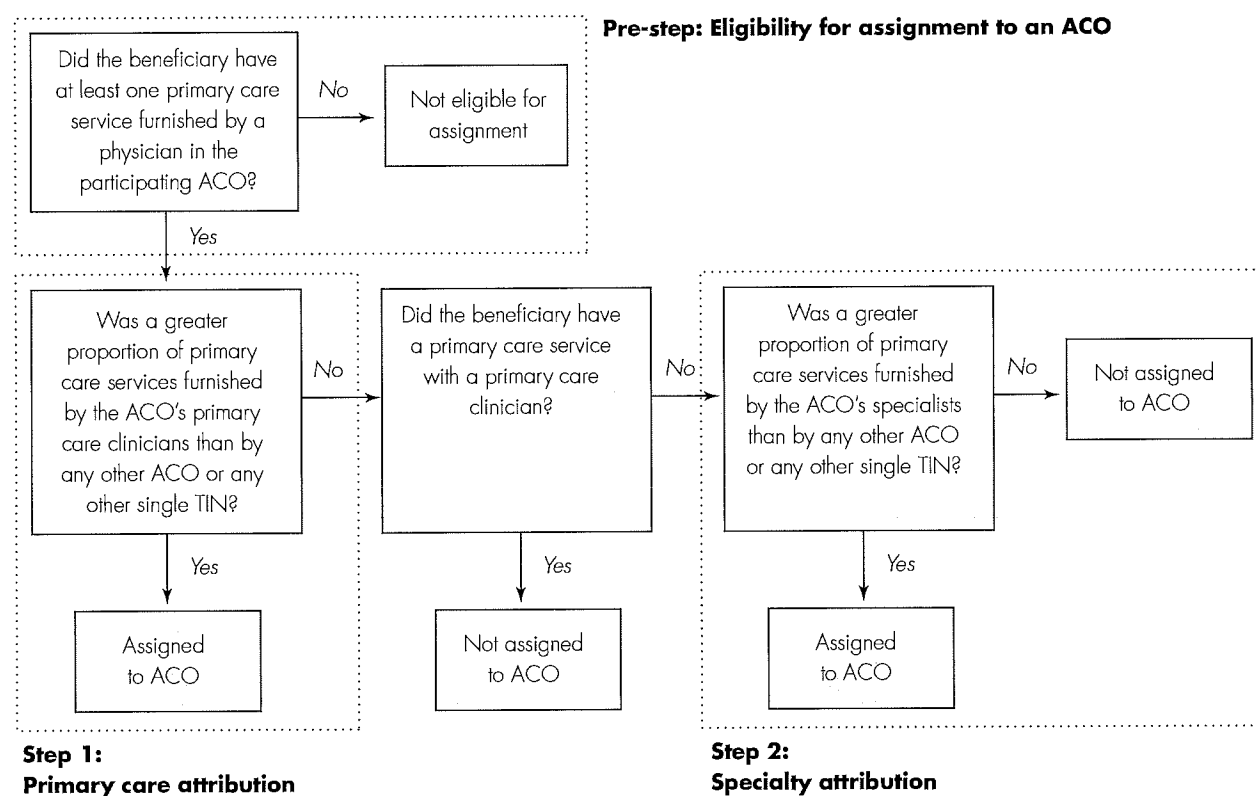
In the Medicare Shared Savings Program (MSSP), beneficiaries are assigned to MSSP accountable care organizations (ACOs) in a multistep process, as shown in Figure 2-1.

In general, the claims history of beneficiaries who are eligible for ACO assignment is reviewed. Beneficiaries are eligible for assignment if they meet certain criteria, including having been in Part A and Part B of Medicare

(continued next page)

**FIGURE
2-1**

ACO assignment in the Medicare Shared Savings Program



Note: ACO (accountable care organization), TIN (taxpayer identification number). According to regulations found in 42 CFR §425.20, an ACO is identified by a Medicare-enrolled TIN that alone or together with one or more other ACO participants constitutes the ACO. The proportion of primary care services is measured by Medicare-allowed charges. Specialty attribution occurs only for beneficiaries who did not have a primary care service with a primary care clinician but did have a service with an ACO specialist.

Source: Centers for Medicare & Medicaid Services. Financial and beneficiary assignment specifications Versions 3–6. <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/program-guidance-and-specifications.html>.

ACOs. There will always be some unwarranted shared savings payments due to random variation, but there could also be unwarranted shared savings payments due to intentional favorable patient selection. For example, if high-cost beneficiaries are disproportionately shifted

out of an ACO in the performance year—but not in the baseline years—performance-year spending will decrease in relation to the ACO's benchmark, which could result in unwarranted shared savings.

Beneficiary assignment in the MSSP (cont.)

for 12 months (so that they have a claims history) and not having been enrolled in Medicare Advantage during that time.

To be assigned to an MSSP ACO, a beneficiary must have at least one primary care service furnished by a physician in the participating ACO. Services are designated primary care services by regulation and must be furnished by an ACO physician in certain

specialties (e.g., family practice, internal medicine, cardiology, endocrinology, gynecology, nephrology, psychiatry, and oncology) but not by a nonphysician. Visits with primary care physicians take precedence in assignment. (More detail on definitions of primary care services and ACO physicians and nonphysicians can be found in online-only Appendix 6-A to the Commission's June 2019 report, available at <http://www.medpac.gov>.) ■

Under Medicare billing rules, providers bill Medicare using taxpayer identification numbers (TINs). TINs can be used to identify the source of Medicare's billings; CMS uses TINs to identify the billings that are associated with each ACO. However, TINs are not unique to each clinician. Rather, a single TIN can comprise a sole physician in one office or a multistate integrated delivery system with many clinicians. Favorable selection of physicians could occur if an ACO stopped providers with high-cost beneficiaries from billing under the ACO's TINs and had those providers bill under a non-ACO TIN. Selection could also occur if an ACO removed just a portion of a high-cost provider's patients from the ACO. The provider could bill for patients with low spending under the ACO's TINs and bill for patients with higher spending relative to their risk score under a non-ACO TIN. While we do not have evidence of widespread patient selection at this time, we did find evidence that ACOs with large shared savings payments benefited from disproportionately high-cost patients being assigned out of their ACO.

An alternative to removing high-cost patients from the ACO would be to retain low-cost patients in the ACO. ACOs appear to achieve this objective through the use of wellness visits. Whether the wellness visits are designed to retain low-cost patients, to improve quality metrics, or to better manage care, the data suggest they result in ACOs achieving a favorable selection of patients, at least when retrospective assignment is used.

ACOs appear to have generated savings for the Medicare program. However, a future risk of provider and patient selection remains. This type of selection can become more problematic if CMS does not address vulnerabilities in the

current system for assigning physicians and beneficiaries to ACOs. Even if a minority of ACOs engage in selection activities, it could diminish the program's ability to generate Medicare program savings in total. For that reason, we investigate how to make the ACO assignment mechanism less susceptible to mismatches between providers' patient spending history used to set spending benchmarks and providers' actual patient spending used to compute ACO spending in performance years.

Use of TINs for assignment in the MSSP raises concerns

To compute MSSP shared savings and losses, CMS compares actual spending for beneficiaries assigned to an ACO with a benchmark that estimates what spending was expected to be for those beneficiaries. To protect both the Medicare program and ACO participants, ACO benchmarks should be computed in a way that most accurately reflects the health care needs of the beneficiaries assigned to an ACO.

Beneficiaries are assigned to an ACO based on a list of TINs that an ACO annually submits to CMS; this collection of TINs represents the clinicians who will be the ACO's participants for the performance year.⁵ As noted above, a single TIN can range from a sole physician in one office to a multistate integrated delivery system with many clinicians (each individual clinician does have a unique national provider identifier (NPI)). To determine the beneficiaries assigned to an ACO, CMS follows a multistep process described in the text box on beneficiary assignment in the MSSP. In short, claims for each beneficiary are grouped by TINs, and if the ACO (defined as a collection

of TINs) provides the plurality of primary care for the beneficiary compared with any other ACO or individual TIN, the beneficiary is assigned to that ACO.

CMS computes an ACO's spending in the baseline years (i.e., the three years before the ACO's first performance year of its MSSP contract) and combines them to create the historical portion of the benchmark.⁶ That historical spending and regional spending are then blended and trended to the performance year to compute the benchmark against which spending in the performance year will be compared. To establish the historical portion of an ACO's benchmark, CMS computes an ACO's historical spending based on the beneficiaries who would have been assigned to the ACO in the ACO's baseline years. Assignment in the baseline years uses the same list of TINs submitted by the ACO for the performance year and uses the same claims-based multistep assignment process shown in Figure 2-1 (p. 26).⁷ (Between baseline and performance years, assigned beneficiaries are not fixed, but TINs are fixed.)

However, the NPIs associated with an ACO's TINs are not fixed—creating a potential mismatch in the calculation of ACO benchmarks. Mismatches of ACO TIN clinicians can occur when NPIs are removed from a TIN, added to a TIN, or associated with more than one TIN—including TINs in a different ACO and TINs outside of an ACO.⁸ We examined the removal of individual primary care physicians (PCPs) (as specified by their NPIs) from TINs participating in the same ACO in 2016 and 2017.⁹ Among the nearly 103,000 TIN–NPI combinations of PCPs in 2016, 7 percent were removed from ACOs in 2017. TIN-level historical benchmarks did not capture the removal of PCPs from these TINs.¹⁰ We also examined the PCP NPIs added to TINs participating in the same ACO in 2016 and 2017. Among TIN–NPI combinations of PCPs in the MSSP in 2017, 29 percent were added to ACO TINs from the previous year. These PCPs were not participants under any of the ACOs' other TINs in 2016. The NPI removals from and additions to TINs capture only the mismatch in TIN clinicians between 2016 and 2017. There was likely a greater mismatch of TIN clinicians between ACOs' performance year and baseline years, which would have spanned at least four years (the performance year and three baseline years). If ACOs manipulate these mismatches to increase the likelihood of receiving shared savings payments without lowering their growth in spending (or avoiding shared losses when increases in spending growth occur), the result creates vulnerabilities in the MSSP.

Three vulnerabilities

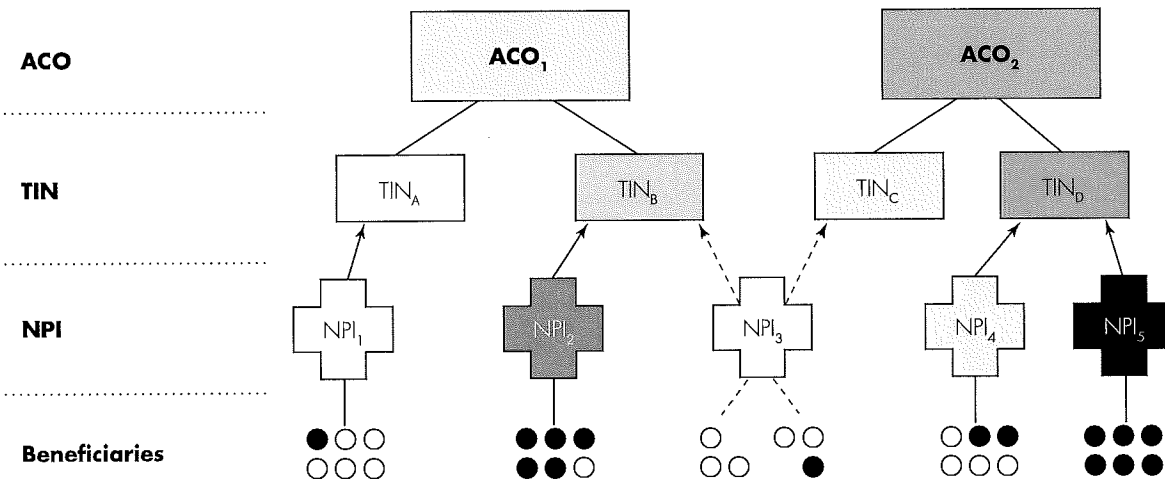
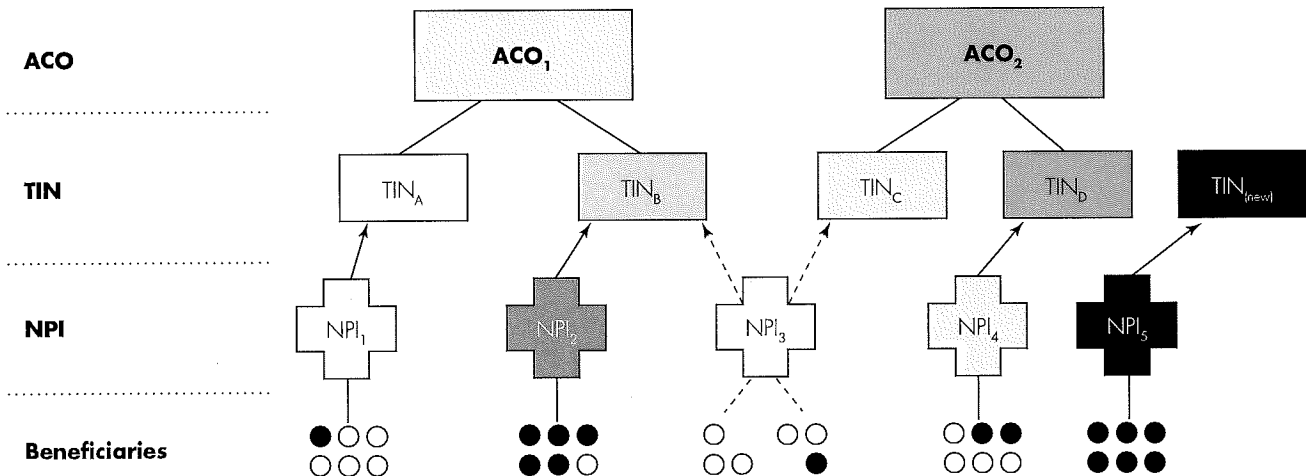
The reliance on TINs to compute the benchmark against which an ACO's financial performance is measured creates three vulnerabilities that could result in unwarranted shared savings.

Clinicians removed from TINs One vulnerability is that an ACO's historical benchmark (based on TINs) is not adjusted when clinicians (and their patients) are removed from its TINs in later years. An ACO could unjustifiably receive “shared savings” by removing high-cost providers from TINs in the ACO. The beneficiaries who would have been assigned to those high-cost providers would remain in an ACO's benchmark but would be removed from the ACO's performance-year spending. The illustration in Figure 2-2 shows this vulnerability in hypothetical ACOs. Before the performance year, ACO₂ removes NPI₅, who has beneficiaries with relatively high spending, from participant TIN_D. The high cost of NPI₅ continues to be in ACO₂'s baseline, which is used to calculate the ACO's benchmark. However, the ACO is not liable for NPI₅ in its performance year, leading to unwarranted shared savings.

The hypothetical example in Figure 2-2 illustrates how the assignment algorithm is vulnerable to shifting the TINs under which NPIs bill. See the text box on anomalous results using TINs (pp. 30–31) for an example of how the current assignment mechanism using TINs could have contributed to some of the anomalous shared savings payments that have been made.

Clinicians added to TINs A second vulnerability resulting from TIN-level benchmarks can occur when providers are added to TINs. In this case, the benchmark may not reflect the historical claims of those providers. In particular, primary care physicians could be added under TINs with which they have no historical claims data (that is, in the baseline years, they billed under a different TIN). An ACO could receive unwarranted shared savings by adding low-cost providers who previously billed Medicare using TINs outside of the ACO's current participant list. The low-cost providers' claims would not be included in the ACO's benchmark calculation but would be included in the ACO's performance-year spending.

Billing high-cost patients under non-ACO TINs A third vulnerability resulting from the use of TIN-level benchmarks is that providers can opt to bill high-cost patients under TINs outside of the ACO's participant list, through referrals or through directly billing to a separate

**FIGURE
2-2****Illustrative example of selection resulting from
changing the TIN under which an NPI bills****Baseline ACO-TIN configuration used for ACO benchmarks****Performance-year ACO-TIN configuration**

Note: ACO (accountable care organization), NPI (national provider identifier), TIN (taxpayer identification number). Each dot represents 1,000 beneficiaries. Black dots represent beneficiaries with relatively high spending; white dots represent relatively low-spending beneficiaries. Lines connect beneficiaries to the NPIs through which their ACO assignment is determined.

Source: MedPAC analysis of Medicare Shared Savings Program assignment algorithm.

Example of anomalous results using identification of ACO participants at the level of taxpayer identification numbers

To illustrate how the movement of providers' national provider identifiers (NPIs) in and out of an accountable care organization's (ACO's) taxpayer identification numbers (TINs) can be associated with anomalous results, we look at an ACO that had large savings relative to its benchmarks in 2016, 2017, and 2018 (Table 2-2). This ACO also exhibited a great deal of volatility in its roster of participating clinicians and the risk profile of its beneficiaries. There is a notable change

in the number and mix of clinicians in the ACO between 2015 and 2016. In those years, the number of primary care physicians declined from 265 to 154, and the number of specialists declined much more, from 565 to 103. This dramatic change in clinicians coincided with the renewal of the ACO's Medicare Shared Savings Program (MSSP) contract. The new contract recalculated the ACO's benchmarks based on beneficiaries assigned to the ACO's TINs from 2013 through 2015.

(continued next page)

**TABLE
2-2**

Example of an ACO with volatile enrollment and clinician participation

	2015	2016	2017	2018
PCPs	265	154	187	240
Specialists	565	103	125	154
Nonphysician providers	89	81	244	294
Assigned beneficiaries	8,597	6,051	5,742	5,451
Risk score	1.35	1.10	1.07	1.06
Benchmark per capita	\$19,859	\$20,720	\$23,181	\$22,929
Spending	\$22,987	\$15,836	\$16,262	\$15,800
Difference	-\$3,127	\$4,884	\$6,919	\$7,130

Note: ACO (accountable care organization), PCP (primary care physician). Shared savings are calculated as a percentage of the difference between the ACO's benchmark and spending. Components may not sum to totals due to rounding.

Source: MedPAC analysis of CMS Medicare Shared Savings public use files.

TIN. At least one ACO reported creating a separate TIN for physicians serving certain high-cost patients to avoid having those patients assigned to the ACO (RAND Corporation 2018). Under these scenarios, more high-cost beneficiaries would be assigned to an ACO's historical benchmark—before providers billed high-cost beneficiaries outside the ACO's TINs—compared with the ACO's performance year.

Use of TIN-NPI combinations for assignment has shortcomings

In the NextGen and Direct Contracting demonstrations, providers are identified at the TIN-NPI level rather than at the TIN level. That approach avoids the problem of the TIN-based benchmarks staying constant even if clinicians are removed. However, benchmarks based on TIN-NPI combinations remain vulnerable to inaccuracies if PCPs are added to ACO TINs. In addition, unlike TIN-based

Example of anomalous results using identification of ACO participants at the level of taxpayer identification numbers (cont.)

At the same time, the number of assigned beneficiaries changed as well, with a 30 percent drop from 2015 to 2016. Many of those beneficiaries had likely been assigned to the ACO through the providers who left. This drop was accompanied by a change in the average risk score for the beneficiaries in the ACO. Between 2015 and 2016, the average risk score dropped from 1.35 to 1.10 and then leveled off in 2017 and 2018.

The 2016 drop in risk score did not correspond with a decrease in the benchmark because the historical spending of beneficiaries assigned to the ACOs TINs did not decrease. The remaining physicians belonged

to TINs with historically high spending and risk scores (from 2013 to 2015) relative to the ACO's performance years (from 2016 to 2018). As a result, the ACO's spending compared with its benchmark switched from being substantially greater than the benchmark in 2015 to being substantially below the benchmark in 2016, 2017, and 2018 (\$4,884 per capita, \$6,919 per capita, and \$7,130 per capita, respectively). After collecting over \$35 million in shared savings from 2016 to 2018, this ACO discontinued its MSSP participation in 2019—when benchmarks would have been based on historical spending from 2016 to 2018. ■

benchmarks, TIN–NPI combination benchmarking would be vulnerable to unwarranted shared savings when an ACO moves an NPI between two of its TINs. In this scenario, an NPI could have spending under one of the ACO's TINs in the baseline years, but that spending would not be captured in the benchmark if the NPI began billing under a new TIN within the same ACO during the performance year. Under TIN–NPI benchmarking, an ACO could unjustifiably receive shared savings by moving low-cost providers between two of its TINs. The low-cost providers would not be in the ACO's benchmark but would be included in the ACO's performance year spending. In the NextGen demonstration, the substantial changing of TIN–NPI combinations between the first and second year of the program prompted methodological changes to how CMS's contractor evaluated the second year of the program. To evaluate quality and spending relative to a comparison group, the evaluator of the NextGen ACO demonstration in its most recent evaluation used NPI-only assignment to create a better match between baseline-year beneficiaries and an ACO's performance-year beneficiaries (NORC at the University of Chicago 2020).

Opportunities to select healthy patients

The savings achieved by ACOs for the program thus far (1 percent or 2 percent) could be vulnerable if ACOs can engage in patient selection that is not reflected in their

benchmarks and subsequently leads to unwarranted shared savings payments. This benchmarking problem could result from having low-cost patients enter into the ACO without changing the benchmark or having high-cost patients exit the ACO without changing the benchmark. We have not seen evidence of pervasive selection thus far, but we are concerned about the incentives as ACO experience matures and shared savings become more reliant on risk adjustment and regional spending.¹¹

One strategy is to use annual wellness visits (AWVs) for assigning patients to an ACO. Patients who have AWVs are generally low cost in the year of the visit. This strategy is easier to pursue under a system of retrospective assignment rather than prospective assignment. Retrospective assignment is technically known as preliminary prospective assignment with retrospective reconciliation. It is also sometimes referred to as concurrent assignment. In its MSSP assignment specifications, CMS most commonly uses the term *retrospective assignment*.

A review of retrospective and prospective assignment

As described earlier, beneficiaries are assigned to an ACO based on which ACO provided the plurality of their qualifying primary care services. Assignment can be based

Retrospective and prospective assignment of beneficiaries to accountable care organizations

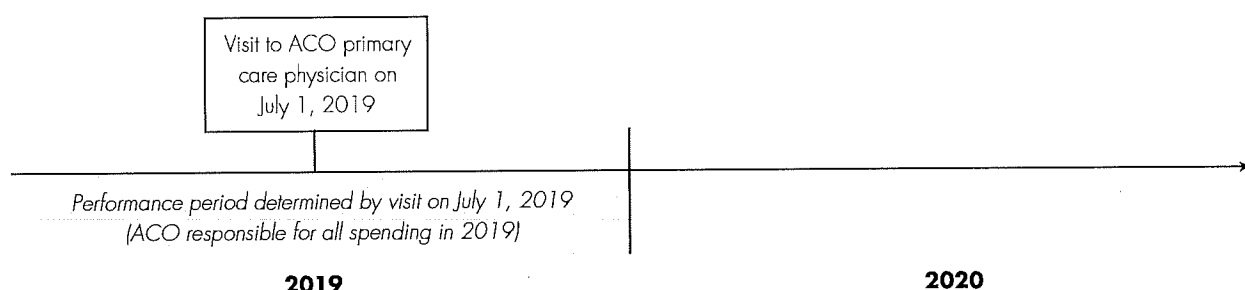
To illustrate the difference between prospective and retrospective assignment, the first two figures show an example of a patient assigned to an accountable care organization (ACO) based on a single primary care visit to an ACO primary care physician on July 1, 2019, under first retrospective and then prospective assignment. As Figure 2-3 shows, under retrospective assignment, the ACO would be responsible for all spending that occurs in 2019, including the six months before the July 1 visit and the six months after the visit, and could include care from non-ACO clinicians in 2019.

Figure 2-4, by contrast, uses the same example of a patient assigned to an ACO based on a single primary care visit to an ACO primary care physician on July 1, 2019, to show that under prospective assignment, the ACO would be responsible for all spending in 2020 (for all applicable months that the beneficiary was in fee-for-service Medicare). All of that care would occur after seeing an ACO clinician, and it could include care from non-ACO clinicians in 2020.

(continued next page)

**FIGURE
2-3**

Performance period under retrospective assignment



Note: ACO (accountable care organization).

on as little as one primary care visit with a physician. Different ACO programs have different rules about which primary care services determine assignment. Most ACOs in the MSSP have used retrospective assignment. Under this approach, a beneficiary is preliminarily assigned to an ACO based on primary care visits during the prior year (e.g., 2018), but the final assignment is determined retrospectively by examining the plurality of primary care visits during the performance year (e.g., 2019). The list of preliminarily assigned beneficiaries will differ from the list of finally assigned beneficiaries to the extent that patients switch clinicians over the two-year period. The difference in assignment lists can be substantial. For example, in

2017, 21 percent of beneficiaries assigned to an ACO preliminarily were not assigned at the end of the year, and 27 percent of those assigned finally were not on the preliminary assignment list.¹²

Under prospective assignment (as used in the NextGen ACO model), beneficiaries' final assignment is made based on their primary care visits during the fiscal year before the performance year.¹³ In other words, under prospective assignment, ACOs know with almost certainty which beneficiaries they are responsible for at the start of the year. By contrast, in retrospective assignment, an ACO ends up responsible for many beneficiaries whom the ACO will not know it is responsible for until well into

Retrospective and prospective assignment of beneficiaries to accountable care organizations (cont.)

If patients see the same primary care physician over multiple years, prospective and retrospective assignment will not differ. However, which assignment mechanism is used has substantial assignment implications for beneficiaries who switch primary care providers from one year to the next. On the one hand, one could argue that it makes sense in the example for the ACO under retrospective assignment to have responsibility for 2019 spending because an ACO physician saw the patient in 2019 and would have some influence over his or her care in the last half of the year. On the other hand, the patient could have had high

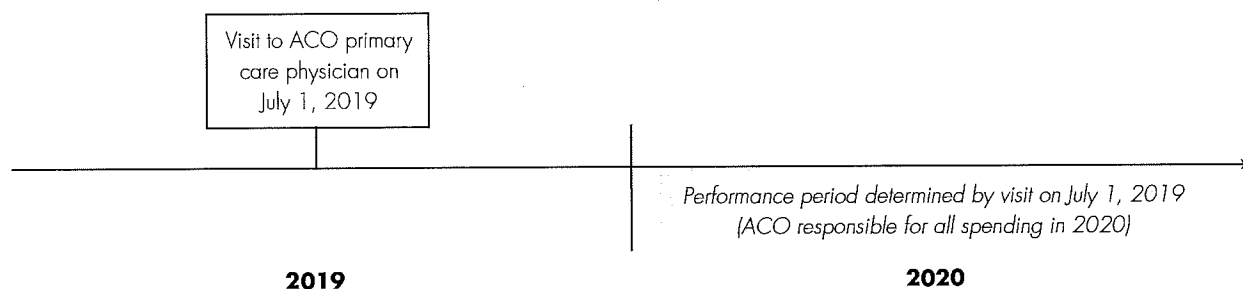
spending before July 1, 2019, and it would be unfair for the ACO to be accountable for spending that occurred before ever seeing the patient.

Under prospective assignment, in which the ACO is responsible for 2020 spending, one could argue that the ACO should have at least a small influence over 2020 spending because it will occur after an ACO physician has seen the patient, and the ACO will receive updates on the beneficiary's health status and medical services received in 2020, even if the beneficiary switches to a physician outside of the ACO.

(continued next page)

**FIGURE
2-4**

Performance period under prospective assignment



Note: ACO (accountable care organization).

the performance year, and the ACO will lose a share of patients it thought it would be responsible for, but is not. The text box on retrospective and prospective assignment (pp. 32–35) illustrates the mechanics of these approaches.

Opportunities to use wellness visits to retain low-spending beneficiaries in ACOs

While patient selection did not appear to have a significant net effect on shared savings in the initial years of the program, patient selection could represent a vulnerability for the ACO program going forward. We first consider

the potential for selection of low-spending beneficiaries in ACOs through AWVs. Currently, ACOs use AWVs more than traditional FFS, AWV patients at their initial AWV tend to have had lower historical spending than other patients, and AWVs have not resulted in Medicare savings. Second, we consider the possibilities for selection against high-spending beneficiaries. The selection of beneficiaries based on their spending patterns could result in unwarranted shared savings payments to ACOs.

Our June 2019 report explored ACOs' use of AWVs and described how AWVs could result in a favorable

Retrospective and prospective assignment of beneficiaries to accountable care organizations (cont.)

Under both retrospective and prospective assignment, the ACO of the physician who saw the patient in the prior year should receive updates on the patient's health status, up until three months after the patient starts to see another physician. In a hypothetical example shown in Figure 2-5, a beneficiary received a September 1, 2019, visit with a primary care provider (PCP) who participates in ACO 1. The patient then has a hospital admission in February 2020 followed by a primary care visit on July 1, 2020, with a different PCP, who participates in ACO 2. In this example, under prospective assignment, ACO 1 would have responsibility for the beneficiary's spending in 2020. Under retrospective assignment, ACO 2 would have

responsibility for the beneficiary's spending in 2020. In both cases, the performance year in question is 2020.

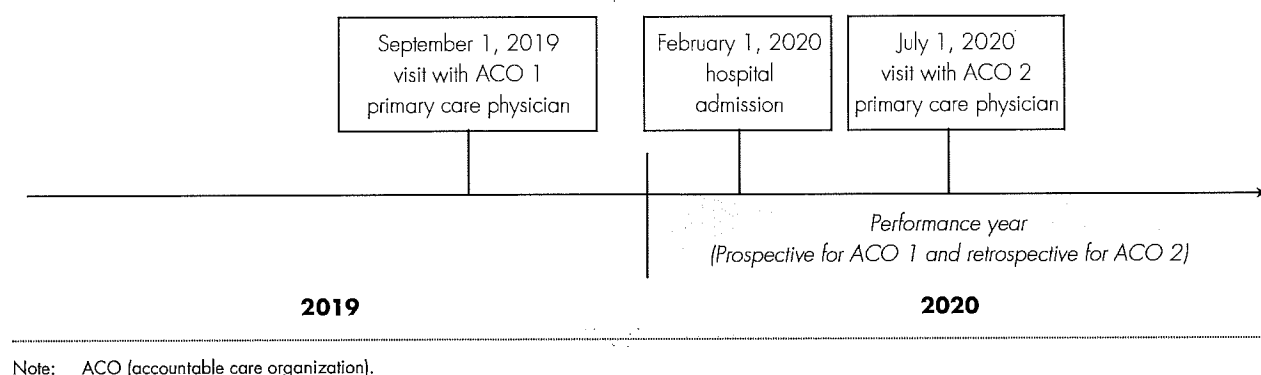
Given this illustrative example of the timing of physician visits, we contrast the Medicare Shared Savings Program's retrospective assignment and information flow with Next Generation ACOs' fully prospective assignment under this scenario (Table 2-3).

The assignment method used can make a difference in which ACO is responsible for a beneficiary's spending in a given year. Under prospective assignment, ACO 1 is responsible for Beneficiary A's spending in 2020; under retrospective assignment, ACO 2 is responsible.

(continued next page)

**FIGURE
2-5**

Assignment under retrospective and prospective assignment for Beneficiary A



selection of patients. While ACOs' motivation for AWVs could be care coordination or improvements on MSSP quality metrics (e.g., to document counseling on smoking cessation or screening for clinical depression), they could also result in keeping relatively healthy beneficiaries assigned to the ACO and receiving higher risk scores from the health risk assessments performed during the wellness visit. In a study of motivations for AWVs, ACOs mentioned patient care needs, performance on quality metrics, assignment, revenue, and Medicare's

hierarchical condition category (HCC) coding of patients (Briggs et al. 2019). However, the study did not find any better performance on cost or quality among ACOs using AWVs as a care management strategy. In addition, the Commission has noted the use of health risk assessments—an essential element of AWVs—to increase HCC scores and has recommended that diagnoses stemming only from these services be excluded from risk score calculations both in FFS and in MA (Medicare Payment Advisory Commission 2016).

Retrospective and prospective assignment of beneficiaries to accountable care organizations (cont.)

There are advantages to prospective assignment. First, under prospective assignment, the ACO that receives information on the patient's health status and health care services at the start of the year will be the ACO responsible at the end of the year. This approach (which mirrors the Medicare Advantage approach) makes population health analytics easier (Table 2-3). Second, prospective assignment makes it easier to construct algorithms to work with other payment policies. For

example, to avoid paying twice for the same savings, CMS would want to know at the beginning of the year whether a patient is in an ACO and not allow that patient to be in a bundled payment initiative in the same year. Making this determination requires prospective assignment so that whether the patient is in an ACO is known with certainty. (An ACO could still initiate its own bundled payment initiative with physicians if it wanted.) ■

**TABLE
2-3**

Information flow under prospective and retrospective assignment

	Under prospective assignment (e.g., NextGen) (Beneficiary A assigned to ACO 1)	Under retrospective reconciliation (e.g., MSSP) (Beneficiary A assigned to ACO 2)
Early January 2020	ACO 1 is told it is responsible for all health care costs in 2020 for Beneficiary A.	ACO 2 receives information on patients it saw in 2019, but receives no information on Beneficiary A because it did not see the beneficiary in 2019.
January	ACO 1 is told about Beneficiary A's historical spending during 2019.	
February 1	If ACO 1 has a relationship with the hospital, the hospital lets ACO 1 know Beneficiary A was admitted.	
April	ACO 1 is told by CMS that Beneficiary A was admitted to the hospital.	
July 1	ACO 1 is initially unaware of the visit to a PCP in ACO 2.	ACO 2 knows that Beneficiary A was seen by one of its doctors and it may be responsible for all costs during 2020.
October	ACO 1 gets an updated report on all spending in the prior quarter including the visit to the PCP in ACO 2 on July 1.	ACO 2 is told by CMS that Beneficiary A may be assigned to it because ACO 2 has the most 2020 PCP-visit allowed charges. ACO 2 first learns about Beneficiary A's 2019 and 2020 spending.
January 2021	ACO 1 is held responsible for all Beneficiary A spending during 2020 (despite being assigned on the basis of a visit in September 2019).	ACO 2 is held responsible for all Beneficiary A spending during 2020 (including during the six months before having any information on the patient).
Note: NextGen (Next Generation), ACO (accountable care organization), MSSP (Medicare Shared Savings Program), PCP (primary care physician).		

The possibility of AWWs resulting in patient selection is particularly concerning in light of patients' relative health status before receiving their initial AWW. We examined the historical risk scores of patients continuously assigned to the same ACO from 2014 to 2016 who had been eligible for ACO assignment since 2012. We compared patients who received their first AWW in 2015 with those who did not.¹⁴ Although both sets of patients were about the same average age (74 years in January 2015), the average risk score of patients who received their initial AWW was relatively lower before receiving the visit. In addition, patients with wellness visits (particularly in the second half of the year) tended to have relatively low spending in the year of the visit, even after adjusting for risk using HCC scores. This finding implies that beneficiaries who are relatively healthy (even adjusting for risk scores) may be more likely to receive wellness visits compared with beneficiaries who need more resource-intensive care.

Support for AWWs is rooted in the assumption that the visits are important elements of care coordination and early intervention that could lead to reduced future spending. However, a November 2019 study found that AWWs did not result in improvements in care or reductions in Medicare spending in FFS from 2008 to 2015 (Ganguli et al. 2019).

It is possible that some ACOs have leveraged AWWs to improve care coordination and patient outcomes. However, the limited evidence thus far suggests that AWWs have had no overall effect on appropriate screening rates, low-value screening rates, referrals for neuropsychiatric and functional issues, emergency department visit rates, or hospitalization rates (Ganguli et al. 2019). While some suggest that AWWs improve patient satisfaction, our beneficiary focus groups suggest that patients have not found AWWs useful for their own care needs. A number of beneficiaries noted the long list of questions that they were asked to answer. Many said they were given the questions in written form, or even online, to fill out before the visit. Some beneficiaries felt that most of the questions did not apply to them. Beneficiaries who spoke favorably of the AWW did not feel the AWW was personally useful to them but spoke of the visit's potential usefulness to high-risk beneficiaries (e.g., those with dementia, home safety issues, or food security issues).

The lack of evidence that AWWs result in Medicare savings exacerbates concerns about their future impact on patient selection and diagnostic coding. The modest savings that ACOs have achieved thus far may have

resulted from care management methods outside of the AWW (e.g., extended office hours) or from eliminating unnecessary care. If most ACOs continue to outpace non-ACO providers in their use of AWWs without any corresponding savings for Medicare or improvement in patient outcomes, the selection of patients through AWWs—even if unintentional—will be an overall vulnerability to the MSSP and could result in unwarranted shared savings.

Opportunities to select against high-spending beneficiaries in ACOs

As with opportunities to select low-spending beneficiaries, there is the potential for selection against high-spending beneficiaries. To determine this potential, we observed characteristics of high-spending beneficiaries that affected their assignment to ACOs and assessed ways the program could be vulnerable to selection against such beneficiaries.

As discussed in our June 2019 report, the assignment of beneficiaries to ACOs and the loss of their assignment often occurs because of changes in beneficiaries' health status; individuals who change health status tend to have rapidly increasing spending compared with those who are continuously assigned (Medicare Payment Advisory Commission 2019). Beneficiaries whose assignments are changed are more likely to have had a hospitalization and use post-acute care during the year their assignment changed. If assignment entry and exit were consistent in the baseline and performance years, such changes would not be an issue. However, if exit of high-spending beneficiaries increases in the performance year and the difference in spending among beneficiaries continuously assigned and those who lose assignment is large, it could improve an ACO's performance relative to its benchmark and lead to unwarranted shared savings.

Techniques to increase the exit of high-spending beneficiaries could include actions at the ACO level, such as moving clinicians with high-spending patients from the ACO to a different TIN, or actions at the PCP level, such as billing those patients under a TIN outside the ACO or counseling patients to seek care elsewhere (presumably from a colleague or other PCP providing care of a similar quality). We found that the shared savings of individual PCPs could be relatively high—providing a material incentive to adjust backroom operations to improve patient selection. We examined earned "shared savings" for each ACO and divided that bonus payment by the number of the ACO's participating PCPs. We found

that 50 ACOs had earned shared savings per PCP of over \$50,000. (The highest was over \$300,000.) Although these ACOs may have used some of the shared savings for ACO administrative costs or shared them with other clinicians, it appears that some ACOs could have had a material incentive to take actions to select against high-spending patients.

Use of NPI for assignment would improve benchmark validity and reduce unintended incentives

Basing benchmarks directly on the individual NPI claims data of an ACO's participating clinicians would be the most accurate method of validly capturing historical spending for purposes of calculating benchmarks and reducing undesirable incentives. Using NPIs' claims data would improve the comparability of beneficiaries assigned in baseline years to those assigned in performance years—reducing opportunities to manipulate shared savings. Because all of an NPI's applicable claims would be used for beneficiary assignment, providers who would be added to or removed from TINs would not affect NPI assignment. Similarly, NPI assignment would not be affected by providers who changed their TIN billing patterns for particular services or beneficiaries. Consequently, the potentially negative incentives associated with TIN-level assignment do not apply to NPI-based assignment.

Implementation of NPI-based assignment for benchmarks could largely follow the same processes as MSSP's TIN-level assignment in which CMS recalculates benchmarks based on an ACO's most recent participant list. Assignment by NPI rather than TIN would not require any change to an ACO's structure, the relationships that ACO clinicians have with other providers, or the billing arrangements of ACO clinicians. MSSP participant lists would continue to consist of TINs (or CMS certification numbers when applicable), but MSSP historical benchmarks would be based on a collection of NPIs that billed to ACO TINs during the performance-year assignment period. All of an NPI's claims in the ACO's market—irrespective of the TIN—would be included in assignment computations. For purposes of calculating benchmarks and performance-year assignment, each clinician's NPI would be associated with only one ACO. For clinicians who bill under TINs spanning multiple

ACOs, the clinician's longest standing participation in an ACO could take precedence. CMS would remove the clinician's NPI from assignment calculations for all other ACOs. Further, assignment would continue to be based on a beneficiary's plurality of primary care visits (using the collection of NPIs that billed under the ACO's TINs during the performance-year assignment period).

Implementing these changes would require that clinicians' claims be used for assignment to only one ACO (providers could continue to see any FFS beneficiary regardless of that beneficiary's ACO assignment or nonassignment). The MSSP currently allows clinicians (through their NPIs) to be listed as participants under TINs in multiple ACOs.¹⁵ Consequently, clinicians with a disproportionately wide range of TIN billing arrangements could be reluctant to participate in the MSSP. Physicians can see patients from multiple ACOs, but if their claims are being used for assignment, their NPI would be used only to assign patients to a single ACO. However, in 2017, 90 percent of ACO assignment was determined by PCP visits, and 95 percent of these clinicians were assigned to one ACO.¹⁶

One potential concern about using NPI-based benchmarking is that ACOs may have more opportunities to engage in within-practice selection—potentially sending beneficiaries with higher needs to clinicians in the same practice who are not part of the ACO but still bill under the same TIN. However, this issue could be addressed by having MSSP participant lists continue to consist of TINs, and require that all NPIs under a TIN in a performance-year assignment period automatically be designated as ACO participants—limiting opportunities for ACOs to benefit from changing the profile of clinicians' patient panels within a practice. Any changes to the case mix between clinicians under the same TIN during the performance year would not reduce the accuracy of the calculation of ACO spending in the baseline years used for the ACO's benchmark.

A second concern about NPI-level assignment relates to movement of clinicians from one geographic area to another. If the clinician joins an ACO or leaves an ACO midway through the performance-year assignment period, his or her Medicare claims history from outside the ACO's market should not be used to compute the ACO's assignment for the performance or baseline years. Doing so would be problematic if the clinician's non-ACO practice area was one with higher or lower payment rates or utilization rates relative to the ACO's market.

**TABLE
2-4****Methods of defining providers for ACO historical benchmarks**

ACO assignment	Current use	Potential inaccuracies	Unintended incentives
Collection of TINs	MSSP	<p>Providers removed from TINs are not accounted for in historical benchmarks.</p> <p>Benchmarks may not reflect the historical claims of providers added to TINs.</p> <p>Providers can use TINs outside an ACO for high-cost beneficiaries without affecting the benchmark.</p>	<p>An ACO could receive unwarranted shared savings by:</p> <ul style="list-style-type: none"> • Removing high-cost providers from TINs. The high-cost providers would remain in an ACO's benchmark but would be removed from performance-year spending. • Adding low-cost providers who previously used TINs outside the ACO. The low-cost providers would not be in the ACO's benchmark but would be included in performance-year spending. • Using TINs outside the ACO for high-cost beneficiaries. High-cost beneficiaries would disproportionately remain in the ACO's benchmark but would not be included in performance-year spending.
Collection of TIN-NPI combinations	Next Generation ACO Model	Providers added to TINs do not necessarily reflect the historical claims of those providers.	An ACO could unjustifiably receive shared savings by adding low-cost providers to TINs. Claims histories of the low-cost providers would not be included in the ACO's benchmark but would be included in performance-year spending.
Collection of NPIs	N/A	When clinicians move from one geographic area to another, they would bring historical spending from their former area unless those claims were excluded.	Physicians used for assignment would have all their patients assigned to a single ACO, meaning that specialists working with two ACOs would have to choose which ACO to assign their patients to in the rare case that the specialist consultation determines assignment.

Note: ACO (accountable care organization), TIN (taxpayer identification number), MSSP (Medicare Shared Savings Program), NPI (national provider identifier), N/A (not applicable). There is no current use of NPI-level historical benchmarks.

Source: Analysis of MSSP provider data and CMS program rules for benchmark calculations.

For example, if a physician moved from San Francisco to Tulsa, CMS would not want to include historical claims from patients who received most of their care in San Francisco when computing assignment for the Tulsa ACO's benchmarks because claims for San Francisco beneficiaries would reflect higher payment rates and different utilization patterns and thus would be a poor predictor of likely spending for similar patients in Tulsa. To address this problem, CMS would base assignment only on claims from within the ACO's (performance-year) market.

Our findings show that the use of NPI data for benchmarks would reduce the potential for unwarranted shared savings and that under TIN-level definitions, changes in the clinicians who make up an ACO's TINs weaken the utility of historical assignment and benchmarks. Table 2-4 is an abbreviated list of the potential methods of and concerns about defining providers when calculating historical benchmarks.

To address (1) the potential mismatch between the clinicians considered in an ACO's baseline years and its

performance years and (2) the incentives to select low-spending patients and exclude high-spending patients, CMS should use NPIs to identify ACO clinicians' claims for assignment in the performance year and those same clinicians' claims for assignment in the baseline year. Properly matching the clinicians included in an ACO's baseline and performance years will allow a more accurate assessment of an ACO's performance and reduce opportunities for unwarranted shared savings.

RECOMMENDATION 2

The Secretary should use the same set of national provider identifiers to compute both performance-year and baseline assignment for accountable care organizations in the Medicare Shared Savings Program.

The set of NPIs used would be those of the clinicians responsible for the ACO's performance-year spending. The recommendation would make the baseline and performance-year spending better reflect the practice patterns of the ACO's performance-year clinicians.

Three corollaries to this recommendation would need to be included:

- If an NPI is used to bill under an ACO's participating TIN during the performance-year assignment period, CMS should use all primary care visits in the ACO's market billed from that NPI (regardless of what TIN the visits are billed under) to assign beneficiaries to that ACO in its performance year and baseline years. Doing so would prevent the ACO from allocating high-spending patients to a TIN not in the ACO. Thus, it would partially address selection against high-spending patients.
- Claims occurring outside the ACO's current market should be removed from assignment calculations to prevent claims from other areas being considered if clinicians either join the ACO after moving from a different market or leave the ACO midway through the performance assignment period and move to a different market.

- Clinicians' claims would be used only for assignment to a single ACO to prevent selection among patients by a clinician billing under multiple TINs.

RATIONALE 2

The integrity of using historical benchmarks requires reliably matching the ACO's performance-year clinicians with the ACO's historical primary care visits. The risk is that allowing ACOs to benefit from changing NPI participation in TINs creates potentially perverse incentives and could produce unwarranted shared savings. ACOs should be rewarded for improving patient outcomes and achieving real savings due to appropriately managing utilization—not for apparent gains that result from mismatches between performance-year and benchmark-year clinicians (whether intentional or unintentional). The recommendation would help reduce unwarranted shared savings by using the same NPIs to compute baseline spending as are used to compute performance-year spending. ACOs that shift providers to TINs outside the ACO would not be able to benefit from a mismatch of NPIs used to create benchmarks and NPIs used to evaluate performance.

IMPLICATIONS 2

Spending

- The recommendation is expected to generate a small reduction in Medicare spending due to reduced shared savings payments. The Congressional Budget Office estimates savings of less than \$50 million over one year and less than \$1 billion over five years.

Beneficiaries and providers

- The recommendation is not expected to affect beneficiaries' care. The recommendation will affect ACOs' shared savings payments only to the extent that ACOs shift NPIs into or out of the TINs under which the ACO submits claims. ■

Endnotes

- 1 In 2019, new ACOs joined the program in July, not January as they had in other years.
- 2 Clinicians with a minimum share of professional services payments (or patients) coming through an A-APM qualify for the 5 percent incentive payment. To qualify for the incentive payment in 2020, for example, clinicians must have received at least 25 percent of their Medicare professional services payments through an A-APM in 2018 or delivered services to at least 20 percent of their patients through an A-APM in 2018. A-APMs include Next Generation ACOs and MSSP ACOs in the highest level of the basic track and in the enhanced track.
- 3 The ACOs we interviewed included physician-led and health system-affiliated ACOs, and the states were in the Southwest, South, and Midwest.
- 4 CVS Caremark has previous partnerships with five other Medicare ACOs through its SilverScript PDP. In 2014, it expanded its ACO collaborations to include an additional seven ACOs (Pioneer and MSSP ACO partners all located in California, Florida, or New Jersey) (Avalere Health 2014).
- 5 In lieu of TINs, the MSSP assigns beneficiaries based on a CMS certification number for ACO participants that are federally qualified health centers, rural health clinics, critical access hospitals, and electing teaching amendment hospitals. For these types of providers in the NextGen ACO demonstration, CMS assigns beneficiaries using a combination of a CMS certification number and a national provider identifier.
- 6 Historical expenditures from the first and second baseline years are trended forward to the third baseline year. Expenditures from the first and second baseline years are also adjusted based on their average risk score differential (represented by a ratio of average risk scores relative to baseline year 3). In computing the historical portion of the benchmark, the third baseline year (most recent) is weighted at 60 percent, the second baseline year is weighed at 30 percent, and the first baseline year is weighted at 10 percent.
- 7 CMS annually recalculates historical benchmarks based on the updated list of TINs submitted by the ACO. The list of participating TINs in each ACO can differ markedly from year to year. We examined the consistency of TINs participating in MSSP ACOs in 2016 and 2017. Among the TINs that were reported as participating in MSSP ACOs in 2016, 15 percent were removed from the ACOs' participant lists in 2017. The share of TINs removed in 2017 was higher for physician-only ACOs (20 percent) than for ACOs with a hospital (12 percent). Among MSSP TINs in 2017, 22 percent were added to ACOs from the previous year. ACOs with a hospital added a slightly greater share of TINs (24 percent) compared with physician-only ACOs (21 percent).
- 8 NPIs included in multiple ACOs also create potential ambiguity in assignment for beneficiaries who voluntarily align themselves with an ACO through their designation of a primary care clinician on the MyMedicare.gov website. At any time during the year, a beneficiary may log into MyMedicare.gov and designate a primary care clinician who they believe is responsible for coordinating their overall care. However, to date, this option has seldom been used by beneficiaries.
- 9 PCPs were identified by specialty codes for general practice, family practice, internal medicine, pediatric medicine, and geriatric medicine. To be eligible for assignment, beneficiaries must have an office visit from at least one of these specialties. The determination of assignment—as measured by the plurality of primary care visits—includes nonphysician providers such as physician assistants and nurse practitioners. However, these providers do not currently report a specialty, which raises some issues such as those who work for an orthopedist being assumed to be providing primary care. The Commission has recommended that these practitioners use their own NPI for billing and report a specialty (Medicare Payment Advisory Commission 2019).
- 10 Among ACOs in the MSSP in 2017, 16 ACOs removed more than 20 percent of the TIN-NPI combinations of PCPs from the previous year.
- 11 For ACOs starting a second MSSP agreement in 2017 or later and for any MSSP ACOs starting any agreement as of July 2019 or later, benchmarks are calculated using a blend of the ACO's own historical spending and the ACO market's regional spending. Each subsequent MSSP agreement requires benchmarks to place greater weighting on regional spending (up to a cap of 50 percent). Before January 2019, ACOs could not increase their risk scores for continuing enrollees beyond the average increase for assignment-eligible beneficiaries with the same demographic characteristics. As of July 2019, ACOs can increase their risk scores by up to 3 percent relative to the assignment-eligible beneficiaries with the same demographic characteristics.
- 12 When examining 2017 preliminary and final assignment, we included only beneficiaries who (1) resided in the same county from 2016 to 2017, (2) did not have any 2017 enrollment in MA, and (3) had at least one month of enrollment in Medicare Part A and Part B in 2017.

- 13 There is a third type of assignment that is partly prospective. Under MSSP prospective assignment, the patient is preliminarily assigned to the ACO based on the prior year's visits. But to maintain that assignment, the patient needs to receive some kind of primary care visit with the ACO (but not necessarily the plurality of visits). Some commercial ACOs apply prospective assignment differently from the NextGen program. For example, under the AQC HMO model in Massachusetts, enrollees pick a primary care physician and then are prospectively assigned based on that choice of primary care physician.
- 14 To compare patients who received their first AWV in 2015 with those who did not, we included only markets where the ACO had at least 100 assigned beneficiaries that received an AWV in 2015. Markets were defined as urban metropolitan statistical areas within a state or all rural counties within a state.
- 15 "Any ACO participant, as identified by the taxpayer identification number (TIN), that has a specialty used in assignment (42 CFR §425.402) and bills Medicare for primary care services must be exclusive to a single Shared Savings Program ACO. However, individual practitioners, identified by individual National Provider Identifiers (NPIs), are free to participate in multiple ACOs if they bill under several different TINs" (<https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/for-providers>).
- 16 PCPs and eight specialties accounted for nearly all MSSP assignment in 2017. Cardiology and hematology accounted for about half of the beneficiaries assigned through specialties.

References

- Avalere Health. 2014. CVS Caremark announces new partnerships with Medicare ACOs to reduce drug spending. <https://avalere.com/insights/cvs-caremark-announces-new-partnerships-with-medicare-acos-to-reduce-drug-spending/>.
- Barnett, M. L., and J. M. McWilliams. 2018. Changes in specialty care use and leakage in Medicare accountable care organizations. *American Journal of Managed Care* 24, no. 5 (May 1): e141–e149.
- Brennan, T. 2017. Evaluating adherence in accountable care organizations. CVS Health commentary. <https://payorsolutions.cvshealth.com/sites/default/files/cvs-health-payor-solutions-evaluating-adherence-in-accountable-care-organizations-commentary-jul-2017-download.pdf>.
- Briggs, A. D. M., T. K. Frazee, A. L. Glick, et al. 2019. How do accountable care organizations deliver preventive care services? A mixed-methods study. *Journal of General Internal Medicine* 34, no. 11 (November): 2451–2459.
- Centers for Medicare & Medicaid Services, Department of Health and Human Services. 2018a. Medicare program: Medicare Shared Savings Program; accountable care organizations—Pathways to Success and extreme and uncontrollable circumstances policies for performance year 2017. Final rule. *Federal Register* 83, no. 249 (December 31): 67816–68082.
- Centers for Medicare & Medicaid Services, Department of Health and Human Services. 2018b. Medicare program; Medicare Shared Savings Program; accountable care organizations—Pathways to Success. Proposed rule. *Federal Register* 83, no. 160 (August 17): 41786–41951.
- Colla, C. H., V. A. Lewis, B. R. Beaulieu-Jones, et al. 2015. Role of pharmacy services in accountable care organizations. *Journal of Managed Care & Specialty Pharmacy* 21, no. 4 (April): 338–344.
- Ganguli, I., J. Souza, J. M. McWilliams, et al. 2019. Association of Medicare's annual wellness visit with cancer screening, referrals, utilization, and spending. *Health Affairs* 38, no. 11 (November): 1927–1935.
- Haber, S., H. Beil, P. Amico, et al. 2018. *Evaluation of the Maryland all-payer model: Third annual report*. Report prepared for the Centers for Medicare & Medicaid Services. Prepared for the Centers for Medicare & Medicaid Services. Waltham, MA: RTI International.
- McWilliams, J. M., M. E. Chernew, B. E. Landon, et al. 2015. Performance differences in year 1 of Pioneer accountable care organizations. *New England Journal of Medicine* 372, no. 20 (May 14): 1927–1936.
- McWilliams, J. M., L. A. Hatfield, M. E. Chernew, et al. 2016a. Early performance of accountable care organizations in Medicare. *New England Journal of Medicine* 374, no. 24 (June 16): 2357–2366.
- McWilliams, J. M., L. A. Hatfield, M. E. Chernew, et al. 2016b. Supplementary appendix to: McWilliams, J. M., L. A. Hatfield, et al. Early performance of accountable care organizations in Medicare. *New England Journal of Medicine* 374: 2357–2366. http://www.nejm.org/doi/suppl/10.1056/NEJMsa1600142/suppl_file/nejmsa1600142_appendix.pdf.
- McWilliams, J. M., L. A. Hatfield, B. E. Landon, et al. 2018. Medicare spending after 3 years of the Medicare Shared Savings Program. *New England Journal of Medicine* 379, no. 12 (September 20): 1139–1149.
- Medicare Payment Advisory Commission. 2019. *Report to the Congress: Medicare and the health care delivery system*. Washington, DC: MedPAC.
- Medicare Payment Advisory Commission. 2016. *Report to the Congress: Medicare payment policy*. Washington, DC: MedPAC.
- Muhlestein, D., W. Bleser, R. Saunders, et al. 2019. Spread of ACOs and value-based payment models in 2019: Gauging the impact of Pathways to Success. *Health Affairs* blog. October 21. <https://www.healthaffairs.org/doi/10.1377/hblog20191020.962600/full/>.
- NORC at the University of Chicago. 2020. *Second evaluation report: Next Generation Accountable Care Organization Model evaluation*. Report prepared by staff from NORC at the University of Chicago for the Center for Medicare & Medicaid Innovation. Bethesda, MD: NORC.
- NORC at the University of Chicago. 2018. *First annual report: Next Generation Accountable Care Organization (NGACO) Model evaluation*. Report prepared by staff from NORC at the University of Chicago for the Center for Medicare & Medicaid Innovation. Bethesda, MD: NORC.
- RAND Corporation. 2018. *Effects of health care payment models on physician practice in the United States: Follow-up study*. Research highlight. Santa Monica, CA: RAND.

Roberts, E. T., J. M. McWilliams, L. A. Hatfield, et al. 2018. Changes in health care use associated with the introduction of hospital global budgets in Maryland. *JAMA Internal Medicine* 178, no. 2 (February 1): 260–268.

RTI International. 2012. *Evaluation of the Medicare Physician Group Practice Demonstration*. Report prepared by RTI International for the Centers for Medicare & Medicaid Services. Baltimore, MD: CMS.

Song, Z. 2020. Personal communication with the author, February 7.

Song, Z., Y. Ji, D. G. Safran, et al. 2019. Health care spending, utilization, and quality 8 years into global payment. *New England Journal of Medicine* 381, no. 3 (July 18): 252–263.

Song, Z., S. Rose, M. E. Chernew, et al. 2017. Lower- versus higher-income populations in the Alternative Quality Contract: Improved quality and similar spending. *Health Affairs* 36, no. 1 (January 1): 74–82.

C H A P T E R

3

**Replacing the
Medicare Advantage
quality bonus program**

R E C O M M E N D A T I O N

- 3** The Congress should replace the current Medicare Advantage (MA) quality bonus program with a new MA value incentive program that:
- scores a small set of population-based measures;
 - evaluates quality at the local market level;
 - uses a peer-grouping mechanism to account for differences in enrollees' social risk factors;
 - establishes a system for distributing rewards with no "cliff" effects; and
 - distributes plan-financed rewards and penalties at a local market level.

COMMISSIONER VOTES: YES 17 • NO 0 • NOT VOTING 0 • ABSENT 0

CHAPTER

3

Replacing the Medicare Advantage quality bonus program

Chapter summary

The Commission maintains that Medicare program payments should take into account the quality of care delivered to beneficiaries. In our June 2018 report to the Congress, we formalized a set of principles for designing Medicare quality incentive programs. Medicare’s quality bonus program (QBP) for assessing quality performance in the Medicare Advantage (MA) program is not consistent with these principles. In our June 2019 report to the Congress, we outlined flaws of the QBP program, which:

- scores too many measures, including “insurance function” or administrative measures;
- uses measures reported at the MA contract level, even for contracts encompassing disparate geographic areas, making plan ratings not necessarily a useful indicator of quality provided in a beneficiary’s local area;
- has allowed companies to consolidate contracts to obtain unwarranted bonuses;
- does not appear to adequately account for differences in enrollee social risk factors;
- has moving performance targets that do not permit plans to know ahead of time how their quality results translate to a QBP score; and

In this chapter

- Quality in Medicare Advantage is difficult to evaluate and the quality bonus program is flawed
- Design of the new MA–VIP addresses flaws in the current MA quality bonus payment system
- Illustrative scoring and payment adjustments under the MA–VIP model
- Replacing the Medicare Advantage quality bonus program with a new value incentive program

- is not budget neutral because it is financed with additional program dollars—unlike quality incentive programs in Medicare’s traditional fee-for-service program that are either budget neutral (balancing penalties and rewards) or penalty only.

The flaws of MA quality measurement must be addressed so that Medicare can have confidence that the MA program encourages and appropriately rewards high quality in a manner that ensures that program dollars are wisely spent.

Fixing MA’s quality incentive program is particularly important. More than one-third of Medicare beneficiaries receive their care through MA plans, and overall program payments in MA totaled about \$274 billion in 2019. In the same year, MA’s QBP cost \$6 billion and is projected by the Congressional Budget Office to cost \$94 billion over 10 years. The Commission has discussed moving Medicare into more value-based payment models in which an entity is accountable for both the cost and quality of care provided to Medicare beneficiaries on a population basis. MA is such a model, but the current state of quality reporting and measurement in MA does not provide a basis for properly evaluating the effectiveness of this model.

In the June 2019 report, we introduced an alternative MA value incentive program (MA–VIP). In this report, the Commission recommends that the Congress replace the QBP with an MA–VIP that includes the five key design elements described below. This recommendation would produce savings for the Medicare program and its beneficiaries. In making this recommendation, which involves a reduction in overall MA payments, the Commission is not rendering a judgment on the appropriate level of aggregate payments to MA plans.

The Commission’s recommended MA–VIP would:

- ***Score a small set of population-based measures.*** The MA–VIP measure set would be tied to clinical outcomes as well as patient/enrollee experience. CMS should develop the MA–VIP measure set through a public review and input process. We anticipate that the MA–VIP measure set would continue to evolve as the quality and completeness of MA encounter data improve and patient-level clinical data from electronic health records and other clinical sources become available for quality measurement.
- ***Evaluate quality at the local market level.*** The MA–VIP would evaluate MA plan quality at the level of local market areas because it provides information about the quality of care delivered in the localities in which beneficiaries seek and receive care.

- ***Use a peer-grouping mechanism to account for differences in enrollees' social risk factors.*** In determining the distribution of quality-based payments in each market area, the MA–VIP would consider differences in plans' enrolled population by stratifying results by defined peer groups, using social risk factors such as eligibility for Medicaid, eligibility for the Part D low-income subsidy, disability status, and area deprivation indexes. Comparing performance among groups with similar characteristics accounts for social risk factors without masking disparities in plan performance, as would be the case if measure results themselves were adjusted by population social risk characteristics.
- ***Establish a system for distributing rewards with no “cliff” effects.*** The MA–VIP would reward or penalize a plan based on the plan's performance relative to other plans in the market using a continuous, prospectively set performance-to-points scale for each measure. The use of continuous performance-to-points scales allows plans that improve to earn points and avoids the “cliff” effect, whereby only those plans achieving a certain level of quality receive bonuses.
- ***Distribute plan-financed rewards and penalties at the local market level.*** The MA–VIP would redistribute a pool of dollars (made up of a share of plan payments within a market area) as rewards and penalties based on a plan's performance compared with the market area's other plans.

To test the proof of concept of the MA–VIP design, we modeled a prototype MA–VIP using currently available data. We calculated quality measure results using administrative data for a set of six measures tied to clinical outcomes, along with patient-reported outcomes and experience measures based on survey data. We modeled the MA–VIP scoring and payment adjustments in 61 local market areas that had at least 3 parent organizations meeting minimum sample size requirements for all measures in our modeling measure set. We used nationally determined performance-to-points scales to convert each parent organization's quality results to MA–VIP points. We accounted for social risk factors in plan populations by stratifying parent organizations' enrollees in each market into two peer groups based on their enrollees' fully dual-eligible status. Each peer group in a market area had a pool of dollars based on 2 percent of the parent organization's payments tied to the peer group.

Overall, our illustrative MA–VIP prototype demonstrates the feasibility of implementing a quality performance measurement program that is consistent with the Commission's principles. In stratifying results by peer groups, the MA–VIP accounts for differences in the social risk factors of plan populations and gives plans the opportunity to earn more rewards for higher quality care provided to their fully

dual-eligible population as compared with their non-fully dual-eligible populations. We found stratifying by social risk factors to produce more fair competition in the majority of markets in our illustrative modeling. We also found that, compared with the QBP, the MA-VIP stratification into peer groups and the market-level comparison approach helps to narrow disparities in payments for plans serving higher shares of fully dual-eligible beneficiaries.

Our results indicated that an MA-VIP was feasible. An illustrative withhold of 2 percent of payments yielded small penalties and rewards for each peer group for most parent organizations in a market area. The magnitude of payment adjustments would change based on the size of the reward pool (our modeling used 2 percent of plan payments, but the percentage could be set higher) and how the performance-to-points scale for each measure is set. Policymakers should consider performance scale methodology and an appropriate amount of payment to fund the reward pool that would drive quality improvement.

The current practice of collecting data and measuring quality at the MA contract level limited the availability of data to use in our modeling, which was conducted at the parent organization and local market level. Moreover, the model is not meant to be an exact formula for how the Congress and CMS should implement an MA-VIP. If a new value incentive program is enacted by the Congress, CMS should use the formal rule-making process to select measures, set performance-to-points scales, define the social risk factors that are accounted for in peer groups, and determine the share of plan payments used to fund reward pools. ■

Quality in Medicare Advantage is difficult to evaluate and the quality bonus program is flawed

The Commission maintains that Medicare payments should not be made without considering the quality of care delivered to beneficiaries and has formalized a set of principles for designing Medicare quality incentive programs (Medicare Payment Advisory Commission 2018a). The Commission has been working to redesign Medicare's range of quality incentive programs to be consistent with these principles, such as with the recommendation to implement a hospital value incentive program (Medicare Payment Advisory Commission 2019c).

Reports by the Commission in 2018 and 2019 discuss at length the difficulties in evaluating the quality of care in Medicare Advantage (MA) (Medicare Payment Advisory Commission 2019c, Medicare Payment Advisory Commission 2018b). Indeed, the state of quality reporting in MA is such that the Commission's yearly updates to MA can no longer provide an accurate description of the quality of care in MA. Also, the current quality bonus program (QBP) is overly complex, distributes financial rewards inequitably, and reports inaccurate information on quality. These flaws must be addressed to ensure that the MA program promotes and appropriately rewards high quality and provides accurate information to beneficiaries and policymakers.

The QBP is costly to Medicare and to taxpayers and beneficiaries who finance the program. The QBP is financed with added program dollars, and the number of entities receiving bonus dollars has increased to the point that the financial incentives of the program no longer achieve the original intention of recognizing only the best performing entities, given that over half of all MA contracts, representing 83 percent of MA enrollment, are in bonus status. The current QBP used trust fund and taxpayer dollars to increase MA payments by about 2.3 percent, or \$6 billion, in 2019. Financing the QBP with additional program dollars is inconsistent with the budget-neutral nature of most fee-for-service (FFS) quality incentive programs (some of which involve only penalties), creating an uneven playing field between MA and FFS (including the quality incentive programs for accountable care organizations (ACOs) in FFS).

Fixing MA's quality measurement and quality incentive program is of the highest importance since more than one-third of beneficiaries receive their care through MA plans, and program payments in MA totaled about \$274 billion in 2019. The Commission has discussed moving Medicare into more value-based payment models in which an entity is accountable for both the cost and quality of care provided to Medicare beneficiaries on a population basis. MA is such a model, but the current state of quality reporting and measurement in MA does not provide a basis for properly evaluating the effectiveness of this model, nor does the current system provide accurate information to beneficiaries. The flaws of MA quality measurement must be addressed so that Medicare can have confidence that the MA program encourages and appropriately rewards high quality in a manner that ensures that program dollars are wisely spent. While the QBP was intended to reward high quality, the QBP has also been the source of added program payments unrelated to quality.

The quality bonus program and its flaws

The Affordable Care Act of 2010 called for CMS to institute a QBP for MA beginning in 2012. The law specifies that a 5-star rating system be used to determine MA plans' eligibility for bonus payments. The statute did not provide additional guidance on the structure or operation of the star system, but CMS had already been using a 5-star rating system to inform beneficiaries of MA quality. Plans rated 4 stars or higher ("in bonus status") are rewarded by receiving an increase in their MA benchmarks of 5 percent or, in some counties, 10 percent. (A higher benchmark can result in an increased level of extra benefits for plan enrollees, but when a benchmark increases because of bonus payments, there is no requirement that all the bonus dollars be used to finance extra benefits. A higher benchmark can also result in a plan increasing its bid—that is, increasing its payments to providers for the Medicare benefit package and retaining more dollars for profit and administration rather than applying the benchmark increase toward the computation of rebate dollars that finance extra benefits.)

MA star ratings are based on 45 measures of clinical quality, patient experience, and administrative performance. For each measure, a contract receives a score from 1 to 5 stars. The categories of measures, as defined by CMS, have different weights: 1 for process measures, 1.5 for access and patient experience measures,

3 for outcome measures, and 5 for the two improvement measures that CMS computes. The overall star rating is the weighted average of all the measures a plan can report (and the plan must report at least half of the measures). Certain adjustments are made to arrive at a final overall star rating, including an adjustment for contracts with high shares of low-income enrollees and enrollees entitled to Medicare on the basis of disability.

For most of the star measures, CMS grades plan performance using a “tournament model” to determine the threshold, or “cut point,” for each level of the star ratings (e.g., the measure value that is the cut point distinguishing between a 4-star and 5-star result for the measure). Under this model, plans are measured against each other’s performance, not against a set performance target. Each year, individual measure results are classified (clustered) into five groups, with the highest group at 5 stars and the lowest at 1 star. Under this system, each of the cut points distinguishing the five groupings can be higher or lower from year to year, thus producing shifting performance targets.

In addition to being the basis of bonus payments, the star rating system is intended to be a source of information about MA quality for beneficiaries (see text box about public reporting of quality information and the MA value incentive program (MA–VIP), p. 59). Star ratings—both the overall ratings and star levels for individual measures—are posted on the Medicare Plan Finder site of Medicare.gov. The ratings are updated each October for the October–December annual election period (when beneficiaries can move among plans or between MA plans and FFS Medicare).

As of February 2020, among MA contracts with any star rating, about 83 percent of MA beneficiaries were enrolled in MA plans in bonus status under the 2020 ratings released in October 2019. We estimate that the QBP constitutes about 2.3 percent of aggregate payments to MA plans, or about \$6 billion a year in additional program costs. This level of additional program expenditure means that all of the nearly 60 million Medicare beneficiaries who have Medicare Part B are obligated to pay an additional \$1 per month in their Part B premium—an obligation that also strains state finances because the states pay the Part B premium for the 12 million Medicare beneficiaries who are dually eligible for Medicare and Medicaid.

The QBP has undergone several changes over the years. Some have been in response to, or consistent with,

recommendations or observations the Commission has made with a view toward improving the QBP. At the same time, policy decisions allowing companies to use the contract consolidation strategy to raise star ratings—by merging lower rated contracts with higher rated contracts and allowing plans to choose the higher rating as applicable to the entire consolidated contract—have been detrimental to the program (Medicare Payment Advisory Commission 2019a).

In addition to concerns about cost, the QBP is flawed in that:

- too many measures are scored, diluting results aimed at assessing quality;
- reporting units do not represent market area performance;
- plans are scored against moving, rather than preset, targets; and
- the QBP’s method of accounting for differences in enrollees’ social risk factors does not appear to be effective at addressing these differences.

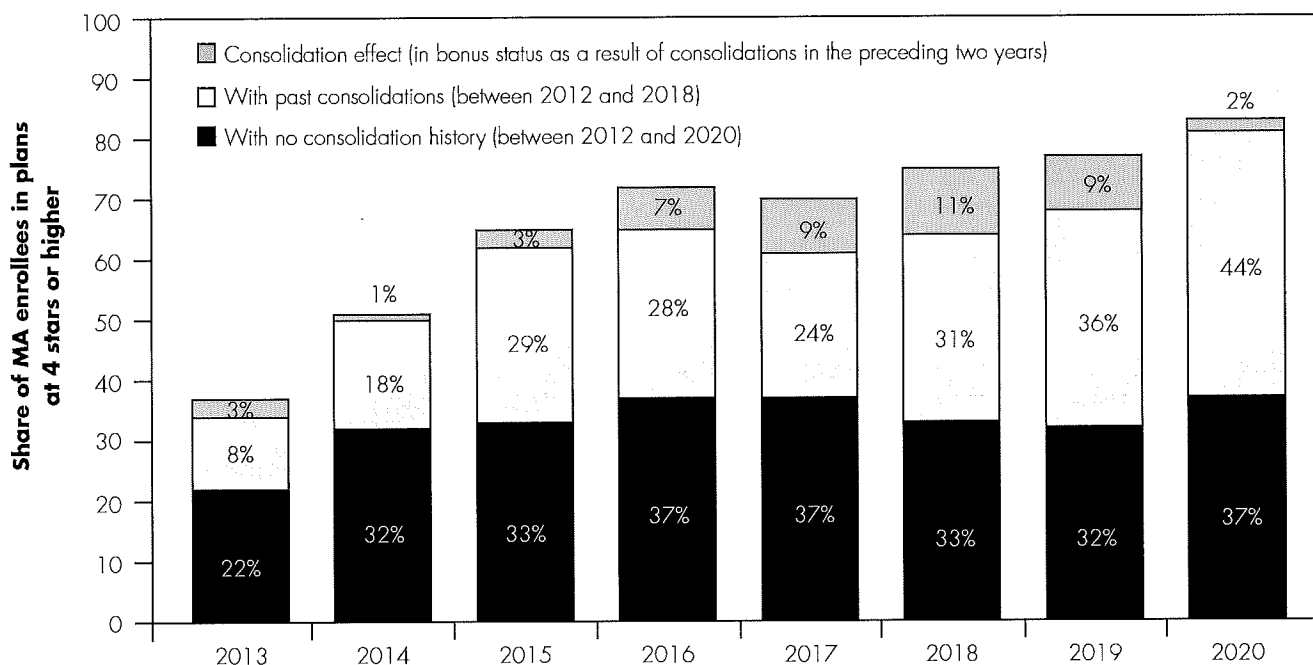
Overpayments in the MA QBP persist as information on quality continues to become less reliable

Both before the QBP and in its early years, very few enrollees were in plans rated 4 stars or higher in CMS’s 5-star system that predates the QBP. In 2011, about 23 percent of MA enrollees were in such plans, and in 2012, the first year of the QBP, about 28 percent of enrollees were in plans meeting the statutory requirement for bonus eligibility (a rating of 4 stars or higher). However, since its inception in 2012, the QBP has been characterized by excess payments unrelated to quality in that CMS used its demonstration authority from 2012 through 2014 to implement an MA-wide demonstration to pay bonuses to contracts rated below 4 stars. Virtually all contracts received bonus payments under the demonstration (e.g., for 90 percent of enrollees in 2012). The Government Accountability Office found that the demonstration resulted in payments of \$8 billion to plans rated below 4 stars (and for payments exceeding other limits the Affordable Care Act of 2010 imposed on QBP payments) and that the demonstration was implemented using questionable legal authority (Government Accountability Office 2012).

In addition, beginning with the March 2015 report to the Congress, each year the Commission has called attention

**FIGURE
3-1**

The share of MA enrollees in plans rated 4 stars or higher increased from about one-third to over 80 percent between 2013 and 2020, with consolidations adding to the share in recent years



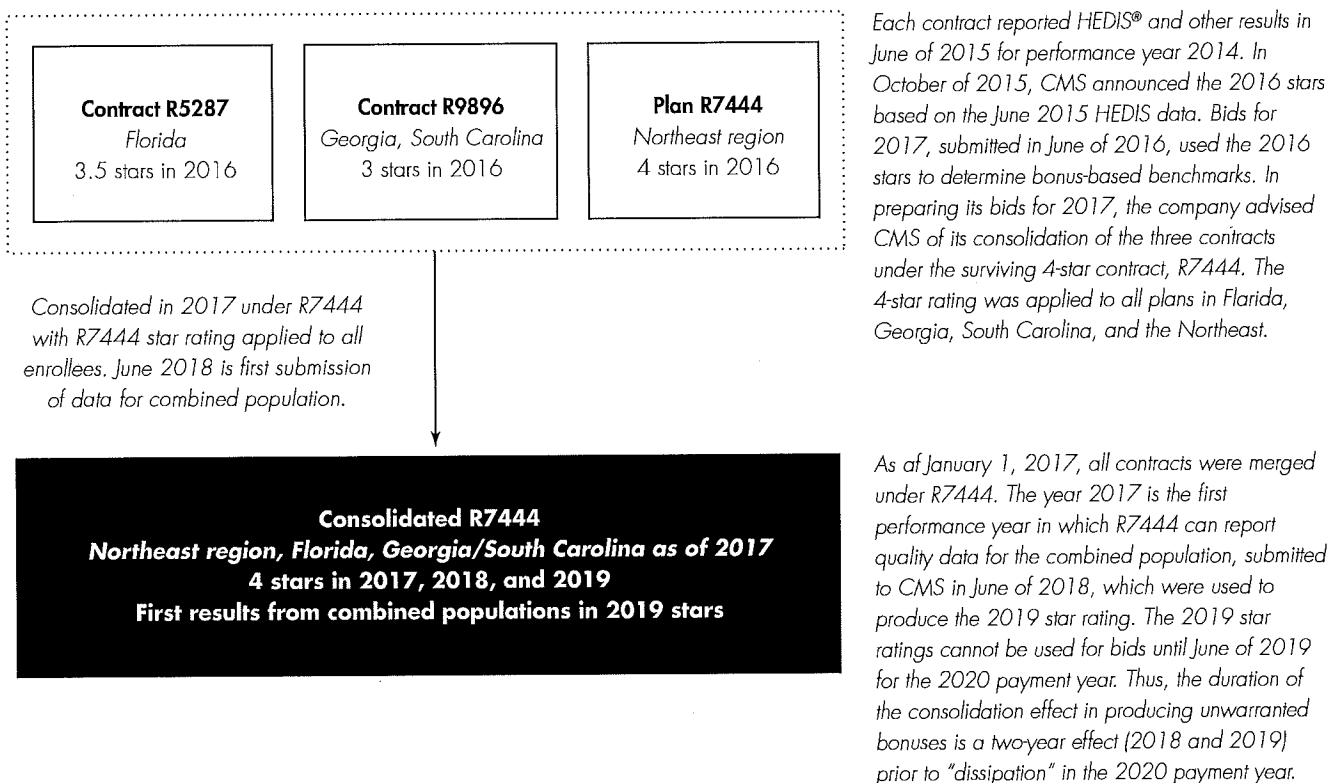
Note: MA (Medicare Advantage). Numbers are rounded amounts. The consolidation effect (marginal consolidation effect) is the share of beneficiaries in bonus-status plans as a result of consolidation, which is the sum of the enrollment moved to bonus status over the two years preceding the year indicated and excluding any enrollees not in a bonus-status plan in the second year of the consolidation effect. It is not until the third year of a consolidation that there can be a star rating that incorporates quality results for enrollees added through a contract consolidation.

Source: MedPAC analysis of CMS enrollment and star rating reports.

to a practice resulting in unwarranted bonuses, which is the use of contract consolidations to achieve bonus status through the mechanism discussed in detail most recently in the March 2018 and March 2019 reports to the Congress (Medicare Payment Advisory Commission 2019c, Medicare Payment Advisory Commission 2018b). Between 2013 and 2020, 81 MA contracts were involved in contract consolidations that moved millions of MA enrollees to bonus-status contracts. Between 2014 and 2018, slightly over 4 million MA enrollees were moved to bonus-status contracts, with plans receiving unwarranted bonuses for those enrollees over at least 2 years (owing to the timing of how star ratings affect payments to plans). In many cases, contracts that were the result of consolidations became absorbed through subsequent consolidations that would maintain the enrollees in bonus-level contracts. By

2020, 83 percent of MA enrollees were in plans with 4 or more stars, up from 33 percent in 2013 (the sum of the three numbers for each year in Figure 3-1). Looking at the shares of 2020 enrollment in any plan with an overall star rating, of the share of enrollees in bonus-level plans (plans rated 4 stars or higher): 37 percent of enrollment is in contracts with no history of any consolidations between 2012 and 2020; 44 percent in contracts that had at least one consolidation between 2012 and 2018; and 2 percent in bonus status as a result of contract consolidations to move to bonus status in the preceding two years (in this case (year 2020), only at the end of 2018, because there was no such consolidation activity at the end of 2019).

Contracts that have had consolidation activity comprise the majority of enrollment in bonus-level contracts (10.7 million of 19.2 million enrollees (56 percent) are in

**FIGURE
3-2****There is a two-year effect on contract bonus ratings after a consolidation and before results for combined populations can be factored into star ratings**

Note: HEDIS® (Healthcare Effectiveness Data and Information Set®). Northeast region for regional plans consists of Connecticut, Massachusetts, Rhode Island, and Vermont.

Source: MedPAC analysis of CMS stars and enrollment data.

contracts with a bonus-level star rating); contracts that include at least one consolidation comprise the majority of enrollment in contracts with any star rating (12.1 million of 23.2 million enrollees, or 52 percent). (Data not shown in Figure 3-1 (p. 53).)

Recent legislation, effective January 1, 2020, changed the policy with respect to consolidations so that consolidated contracts receive the weighted average star rating of the combined contracts. The new policy still permits organizations to obtain unwarranted bonuses by combining lower rated contracts with higher rated contracts when the averaging method yields an overall bonus-level star rating. The legislation has thus narrowed, but not eliminated, the

opportunities for plans to obtain unwarranted bonuses through consolidations.¹

Plans are also employing other strategies to obtain unwarranted bonuses. One strategy capitalizes on the CMS policy that gives new contracts under an existing parent organization the average star rating of the parent organization. In one instance, a company started a new contract as of January 1, 2020, but was able to move more than 100,000 enrollees from counties where it terminated a prior contract into the new contract. The new contract will have a 4-star rating for bidding purposes for the 2021 payment year (2020 bids) as well as for the 2022 payment year (if the company maintains a 4-star average) because

the new contract will not receive its own star rating until October 2021—too late to use for the June 2021 bidding that affects payments in 2022. Medicare beneficiaries will not see star ratings for new contracts until at least two years after the inception of the contract.

CMS has also permitted a company to deconsolidate a set of regional preferred provider organization (PPO) contracts after a consolidation that allowed the company to receive unwarranted bonuses (Medicare Payment Advisory Commission 2017). By restoring the preconsolidation contract configuration, the company is likely to have at least one contract in bonus status while the consolidated configuration would have been a nonbonus situation for all enrollees. The option of deconsolidation after a consolidation, and the ability to change from a consolidated to a deconsolidated configuration (or vice versa) from one year to the next—particularly if the option extends to local contracts as well as regional contracts—is thus another strategy that can result in unwarranted bonus payments.

The lasting effects of consolidations

In addition to being the source of unwarranted bonus payments, past consolidations have produced large multistate contracts, resulting in beneficiaries receiving inaccurate information about MA quality in their local market area. The detrimental effect of past consolidations on the accuracy of plan information about quality cannot be undone. As we have noted, more than half of all MA enrollees are in plans in which the star ratings and quality data reported at the Medicare.gov website are unlikely to accurately reflect the local quality of care. While the recent legislation lessens the concern over unwarranted bonus payments, the continuing ability of plans to consolidate has the potential to exacerbate the information vacuum that beneficiaries have faced because of past consolidations. In addition, CMS continues to permit contracts with wide, disparate geographic areas for new contracts, which perpetuates the problem.

When does the consolidation effect dissipate?

An issue that the Commission has discussed revolves around the estimate of the program expenditures for the bonus program and whether the figure of approximately \$6 billion annually will be less in future years as the effect of consolidations on star ratings dissipates. To be clear about what the \$6 billion represents, it is the total program cost of the QBP—not solely the dollars expended for

unwarranted bonuses. The assumption of a dissipation effect is that the total program cost of the QBP will be less in the future because the star rating of the consolidated organization will decline once the rating is determined based on results for the combined set of enrollees. Figure 3-2 shows how the effect of a consolidation on star ratings will manifest after two years in a specific case.

For all plans, the 2020 star ratings, affecting 2021 payments, are based on performance in 2018 (for Healthcare Effectiveness Data and Information Set[®] (HEDIS[®]) measures) and therefore do not reflect any effects from consolidations occurring before 2018.² Because there was no consolidation activity at the end of 2019, the consolidation effect on 2020 stars is composed entirely of consolidations at the end of 2018. The 2018 consolidations affected 9 contracts with about 550,000 enrollees, and the total number of enrollees after consolidations, in the remaining 6 combined contracts, was a little over a million enrollees. It appears that of the 1 million enrollees in this set of beneficiaries, about 380,000 will be in contracts with a star rating below 4 stars, based on the computation of a weighted average of the last known ratings of the individual contracts. Therefore, about 2 percent of all enrollees in bonus-level plans reflects the effect of consolidations on 2020 star ratings. Thus, the potential for future dissipation of the consolidation effect is of limited magnitude and will not materially reduce the number of enrollees in bonus-level plans.

Design of the new MA-VIP addresses flaws in the current MA quality bonus payment system

In the June 2019 report to the Congress, we described an alternative to the QBP. The MA-VIP is designed to be patient oriented, encourage coordination across providers and time, and promote delivery system change but not be financed with added program dollars (consistent with the Commission's original conception of a quality incentive program for MA). The MA-VIP to replace the QBP would:

- score a small set of population-based measures,
- evaluate quality at the local market level,
- use a peer-grouping mechanism to account for differences in enrollees' social risk factors,

**TABLE
3-1****How the proposed MA-VIP design addresses flaws in the current MA quality bonus program**

Issue	How addressed in the MA-VIP
<i>Too many measures, not focused on outcomes and patient/enrollee experience:</i> The QBP adjusts payment based on plan performance on more than 40 measures that include process and insurance function measures. Many measures are collected through sample medical record reviews.	<i>Score a small set of population-based measures:</i> The MA-VIP adjusts plan payment based on plan performance on a small set of measures tied to clinical outcomes as well as patient/enrollee experience measures.
<i>Contract-level quality measurement is too broad and inconsistent:</i> Contracts can encompass broad, noncontiguous areas, and companies have had financial incentives to create larger multistate contracts. Contract-level reporting does not provide an accurate picture of quality for many areas.	<i>Evaluate quality at the local market level:</i> Evaluation of quality is at the local market level and no longer determined at the contract level.
<i>Ineffective accounting for social risk factors:</i> It is not clear that the current MA peer-grouping mechanisms are effective. Plans serving high-needs populations are less likely to receive bonus payments.	<i>Use a peer-grouping mechanism to account for differences in enrollees' social risk factors:</i> The MA-VIP stratifies enrollees into peer groups based on social risk factors and then calculates quality scores for each peer group.
<i>"Cliff" effect system of awarding bonuses in which only plans receiving a set rating receive bonuses:</i> The QBP scoring has a cliff effect, whereby only those contracts at or above a 4-star overall average receive bonuses.	<i>Establish a system for distributing rewards with no "cliff" effects:</i> The MA-VIP scores plan quality measure results against a continuous, performance-to-points scale that is known ahead of time.
<i>Bonus financing is reward only:</i> With financing from additional program dollars, the QBP is inconsistent with the budget-neutral FFS quality incentive programs and inconsistent with the Commission's original conception of a quality incentive system for MA plans.	<i>Distribute plan-financed rewards and penalties at local market level:</i> The MA-VIP redistributes a pool of dollars (made up of a share of plan payments) as rewards and penalties based on a plan's performance compared with the market area's other plans.
Note: MA-VIP (Medicare Advantage value incentive program), QBP (quality bonus program), FFS (fee-for-service).	

- establish a system for distributing rewards with no "cliff" effects, and
- distribute plan-financed rewards and penalties at a local market level.

Table 3-1 summarizes the MA-VIP design and how it would address the QBP's design flaws.

Ideally, an evaluation of quality in MA would be based in part on a comparison with the quality of care in traditional FFS Medicare, including accountable care organizations, in local market areas (Medicare Payment Advisory Commission 2010a). Some research suggests that MA does have better quality, but a definitive finding is not possible because data sources for comparing MA with traditional FFS at the local market level are limited

(Medicare Payment Advisory Commission 2019c). Therefore, our proposed MA-VIP design does not yet include a component for FFS comparison. In the future, better encounter data from MA and expanded patient experience and patient-reported outcome surveys will help enable comparisons of the two programs.

Score a small set of population-based measures

Over the past several years, the Commission has expressed concern that the QBP is "overbuilt," by including "insurance function" or administrative measures and by relying on many clinical process measures that are weakly correlated with health outcomes of importance to beneficiaries and the program. The majority (31 of the

45 measures in the 2020 star ratings) are either process measures loosely tied to clinical outcomes (e.g., adult body mass index assessment, which simply indicates whether a person's body mass index was recorded in the medical record) or administrative measures (e.g., call center foreign language interpreter and TTY, or teletypewriter, availability). Many are plan-reported measures and require medical record review from a sample of enrollees. The proliferation of measures that are scored in the QBP gives plans several avenues to achieve a bonus-level overall rating, even if their performance is uneven and their results for outcome measures are below bonus-level performance.

Among the Commission's principles for Medicare quality incentive programs is the need to include a small set of population-based measures tied to clinical outcomes as well as patient/enrollee experience. Table 3-2 (p. 58) presents an illustrative example of an MA-VIP measure set consistent with this principle.³ The set includes measures that plans can influence through access to evidence-based clinical care, care coordination, and medication reconciliation. This illustration is not intended to be a definitive list, and CMS should develop the MA-VIP measure set through a public review and input process.⁴

The illustrative set of MA-VIP measures does not include many of the process measures and insurance function measures that are currently scored in the QBP, under the rationale that health plans should be held accountable for their insurance functions through compliance standards and enforcement and through public reporting, not through a quality payment program. Outside of the MA-VIP measures tied to payment, Medicare can use other quality measures and compliance standards to monitor MA plan performance and publicly report this information to encourage improvement (e.g., star ratings and display measures) (see text box on public reporting of quality information and the MA value incentive program, p. 59). For example, Medicare can continue to collect, track, and publicly report plan disenrollment rates.

So that the MA-VIP measures are not unduly burdensome for plans and providers, they should generally be calculated or administered by CMS, preferably with data that are already reported, such as claims, encounters, and enrollee survey data. In November 2019, the Commission discussed the importance of including in the MA-VIP measure set a small number of prevention and chronic care

management measures that are tied to clinical outcomes. Because of the lack of clinical information currently available in administrative data, plans would need to continue to gather data (e.g., hemoglobin A1c lab results for diabetic patients) from a sample of enrollee medical records and report validated measure results to CMS for some of the measures (for example, the HEDIS measures).

The MA-VIP measure set should evolve as better data and measures (e.g., lung cancer screening, patient-reported outcomes for depression and musculoskeletal conditions) become available. As MA plans continue to report encounter data to CMS for risk adjustment and other purposes, the completeness of the encounter data—specifically outpatient encounter data—may improve. Also, measure developers are beginning to produce specifications for plans to calculate measure results using data outside of traditional administrative (claims/encounter) data. The National Committee for Quality Assurance recently published measure specifications for health plans to calculate a small number of HEDIS measures using electronic clinical data systems, such as electronic health records, immunization information systems, and disease/case management registries.⁵ These digital measures have the potential to reduce plan and provider burden in collecting measure results and for plans to calculate measure results on the entire plan population as opposed to a sample of patient/enrollee medical records. However, these digital measure specifications are early in development and implementation and thus would not be available for scoring in the MA-VIP in the near future.

The illustrative MA-VIP measure set covers five measure domains (or measure groupings): (1) ambulatory care-sensitive (ACS) hospitalizations, (2) readmissions, (3) patient-reported outcomes, (4) patient/enrollee experience, and (5) staying healthy and managing long-term conditions. The five domains are generally consistent with the MA star rating domains. We assume that, like the star rating measure set, CMS would seek public input in developing the domains and that weighting of those domains would take into account interests shared by the Medicare program and its beneficiaries. When determining a star rating for each domain, CMS currently weights outcome and patient experience measures more than process measures.

The illustrative measure set includes 12 measures across the 5 domains (Table 3-2, p. 58), focusing on measures

**TABLE
3-2****Illustrative MA-VIP measure set tied to
clinical outcomes and patient/enrollee experience**

Domain	Measures	Data source used to calculate measure results
ACS hospital use	1. ACS hospitalizations 2. ACS emergency department visits	Administrative data
Readmissions	3. Risk-adjusted rate of unplanned readmissions	Administrative data
Patient-reported outcomes	4. Improved or maintained physical health status 5. Improved or maintained mental health status	HOS survey data
Patient/enrollee experience	6. Getting needed care 7. Rating of health plan	CAHPS® survey data
Staying healthy and managing long-term conditions	8. Annual flu vaccine 9. Breast cancer screening 10. Colorectal cancer screening 11. Controlling high blood pressure 12. Diabetes: hemoglobin A1c poor control	CAHPS survey data, administrative data, medical record review

Note: MA-VIP (Medicare Advantage value incentive program), ACS (ambulatory care-sensitive), HOS (Health Outcomes Survey), CAHPS® (Consumer Assessment of Healthcare Providers and Systems®). Data sources used to calculate quality measure results include administrative (claims, encounter) data, information from medical record review, and survey data. Consumer Assessment of Healthcare Providers and Systems® is a registered trademark of the Agency for Healthcare Research and Quality.

that are patient oriented and that encourage coordination and promote delivery system changes. They are also closely tied to clinical outcomes and patient/member experience.

One important note about the illustrative measure set is that it would allow the Medicare program to compare MA plan quality within and across market areas, but would not allow a comparison of FFS and MA plan quality in market areas, which is an ultimate goal for the Medicare program. Such a comparison is not possible mainly because some of the measures (e.g., controlling high blood pressure) require MA plans to use clinical data to calculate results, and the Medicare program cannot currently access this level of clinical information from FFS providers. Also, CMS currently collects FFS Consumer Assessment of Healthcare Providers and Systems® (CAHPS®) results and reports them at a state level (and substate level for larger states) and not market-area level.⁶ CMS no longer fields the Health Outcomes Survey (HOS) for the FFS population. The Commission recommended in 2010 that the Secretary collect and

report HOS data for the FFS population (Medicare Payment Advisory Commission 2010b).

The MA-VIP illustrative measure set includes the following:

- **ACS hospital use:** Hospitalizations and emergency department (ED) visits due to ACS conditions such as diabetes and pneumonia are potentially preventable if ambulatory care is provided in a timely and effective manner. Patients may have required acute-level services at the time they sought care, but the need for the admission or ED visit might have been avoided with appropriate ambulatory care and coordination activities. Rates of ACS hospitalizations and ED visits can reflect an MA plan's quality of care because high-quality MA plans should be able to manage beneficiary, hospital, and physician relations to coordinate care and provide appropriate access (Wholey et al. 2003). In practice, not every ACS hospitalization or ED visit can be avoided, but risk-standardized rates can reveal relative quality.

Public reporting of quality information should complement the MA value incentive program

CMS annually calculates the Medicare Part C (Medicare Advantage (MA)) and Part D star ratings to represent the quality of health and drug services received by beneficiaries enrolled in MA and in prescription drug plans (MA prescription drug plans and stand-alone prescription drug plans, or Part D plans). CMS publishes overall ratings for contracts, consisting of 1 to 5 stars (5 is the highest rating), on the Medicare Plan Finder website for each MA plan available to beneficiaries. On the Plan Detail web pages, consumers (i.e., beneficiaries, family members, counselors, brokers) have the option to view more about a health plan's and drug plan's quality information, including domain summary star ratings such as Staying Healthy, Managing Chronic Conditions, and Member Experience, as well as star ratings for the individual measures that make up each domain. CMS also reports some newer measure results that are not part of the star-rating calculations.

There are two main objectives for publicly reporting Medicare quality information. The first is to increase the accountability of health care organizations and providers, which offers patients, payers, and purchasers a more informed basis on which to hold providers accountable (e.g., directly through purchasing and treatment decisions). The second objective is to maintain standards and stimulate improvements in the quality of care through economic competition (reputation and increased market share) and by appeals

to health care professionals' desire to do a good job (Marshall et al. 2003). Researchers have identified and tested best practices on how to display comparative information to best meet the objectives of public reporting. Many such practices are incorporated in the MA star ratings—for example, using only a small number of data points (or the single data point of an overall star rating), with more detailed information available in a second or even third layer for those who want it (Agency for Healthcare Quality and Research 2020, Aligning Forces for Quality 2009).

Concurrent with the MA value incentive program's direct financial incentive for MA plans to improve care, CMS should continue to have a system and vehicle for publicly reporting quality information to beneficiaries. The design elements of both the quality payment and public reporting programs should generally align. For example, the local market area unit of measurement provides a more accurate picture of quality both for financially rewarding or penalizing performance and for informing beneficiary choice. Medicare should tie performance-based payment to a small set of measures linked to outcomes, but public reporting could include additional measure results to hold MA plans accountable for those measures. What quality information to report and how to report that information to consumers is a separate program design question that should be informed by research, best practices, and stakeholder input. ■

- **Readmissions:** Hospital readmissions are disruptive to patients and caregivers and costly to the health care system; they also put patients at additional risk of hospital-acquired infections and complications. Measuring and adjusting payments based on a plan's readmission rates holds the plan accountable for ensuring that beneficiaries have the discharge information they need and encourages the plan to facilitate coordination with other providers.
- **Patient-reported outcomes:** Beneficiaries are a valuable source of information on outcomes, so the MA–VIP should include enrollee-reported outcomes to assess the quality of care MA enrollees receive. MA plans are required to collect HOS results from a random sample of their Medicare enrollees and, two years later, to survey the same beneficiaries again (if they are still enrolled in the plan). Because the HOS often produces results showing no significant

outcome differences among MA plans, we encourage CMS to continue to improve the HOS instrument to meaningfully capture patient-reported outcomes, for example, by revising the number of surveys required to calculate reliable results (Medicare Payment Advisory Commission 2010b, Rose et al. 2019, Safran 2019).⁷

- **Patient/enrollee experience:** The MA–CAHPS is a national standardized survey instrument and data collection method for measuring enrollees’ perspectives on the quality of health services provided by MA plans. The survey results are used to calculate seven core measures of enrollee experience; they are star measures in the QBP and are publicly reported on the Medicare Plan Finder website, but the MA–VIP could score a subset of these measures, such as the measure for getting needed care and enrollees’ rating of their health plan.
- **Staying healthy and managing long-term conditions:** Preventive services, such as cancer screenings, are an important aspect of health care because they help beneficiaries stay healthier and get more-effective treatment. Chronic disease management is essential to both improving individuals’ health outcomes and potentially containing costs for the Medicare program. MA plans have multiple mechanisms (e.g., clinician incentives, case management, beneficiary screening reminders) to improve the preventive care and chronic care management their enrollees receive, so related measures tied to clinical outcomes should be included in the MA–VIP. These related measures include annual flu vaccine, breast cancer screening, colorectal cancer screening, controlling high blood pressure, and monitoring and controlling diabetes.

Evaluate quality at the local market level

The Commission has a long-standing recommendation that Medicare collect, calculate, and report quality measurement results in MA at a geographically local level because of differences in quality across geographic areas (Medicare Payment Advisory Commission 2010b). A major reason for the flaws in the current QBP is that the unit of measurement for evaluating and reporting on quality is the MA contract, yet MA contracts can cover disparate geographic areas. For example, one insurance company was allowed to have a contract with a service area consisting of counties in Hawaii and Iowa. The star rating for this contract would reflect performance

in two completely different service areas and may not accurately reflect plan quality in either area—making it impossible for the Medicare program to evaluate quality and for beneficiaries in these areas to reliably compare the quality of care when choosing an MA plan. As previously discussed, for 2020, CMS has permitted a number of new multistate contracts covering noncontiguous states. Another problem with using contract-level quality measures is that MA organizations can consolidate contracts, as discussed in an earlier section of the chapter.

We calculated quality results for the illustrative MA–VIP model, looking at each parent organization as identified in CMS data (e.g., United, Aetna, Kaiser Foundation Health Plans, Anthem) within a local market area (e.g., Washington, DC) rather than at the contract level.⁸ We included all the parent organization’s MA products (e.g., HMOs, PPOs, special needs plans) in the local market area quality results. Measuring at the product-type level would likely be too narrow for calculating results; measuring at the level of the parent organization is preferable because provider networks are substantially similar across product types, and Medicare should have the same expectations across all MA products.

Use a peer-grouping mechanism to account for differences in enrollees’ social risk factors

In evaluating quality, Medicare should consider, as necessary, differences in enrollee populations, including social risk factors. Medicare should stratify plan enrollment into groups of beneficiaries with similar social risk factors to determine payment adjustments. Comparing performance among groups with similar characteristics accounts for social risk factors without masking disparities in plan performance, as would be the case if measure results themselves were adjusted by population characteristics. (Outcome measures can be adjusted for patient-level clinical factors such as age, sex, and comorbidities.)

Currently, the QBP takes into account differences in a plan’s enrolled population, including social risk factors, by adjusting overall star ratings. CMS instituted a type of peer-grouping mechanism that modestly adjusts a contract’s overall star rating based on a contract’s share of low-income and disabled enrollees. Nevertheless, in our June 2019 report to the Congress, we showed that plans with a higher proportion of lower income enrollees continue to have lower overall star ratings.

We propose calculating the MA–VIP within a local market area with stratified quality scores for fully dual-eligible enrollees (Peer Group 1) and all other enrollees (Peer Group 2).⁹ In our illustrative MA–VIP model, we use eligibility for full Medicaid benefits (a Medicare beneficiary’s “dual eligibility”), as we do in the hospital value incentive program (HVIP), as a proxy for whether a plan’s enrollees are more difficult to treat. Individuals with full Medicaid benefits are much more likely than other Medicare beneficiaries to be disabled, have multiple chronic conditions, and have functional impairments. Policymakers could consider using other social risk factors to define peer groups, such as beneficiaries qualifying for the Part D low-income subsidy, disability status (which is a current adjustment factor in the MA QBP), and area deprivation indexes, with the definitions subject to refinement as more data became available. When determining the number of peer groups, policymakers will need to weigh the reporting burden (e.g., collecting a reliable sample of patient experience surveys for each group) and the ability to calculate valid measure results for smaller populations.

Establish a system for distributing rewards with no “cliff” effects

The Commission holds that Medicare quality programs should give rewards based on clear and absolute performance targets. However, as currently implemented, MA’s QBP bases bonuses solely on a comparison of results achieved among plans in each year—regardless of overall trends in performance and without assessing whether there should be an expected minimum level of performance for bonus eligibility. Plans do not know in advance whether a certain level of performance is or is not bonus-level performance for a given measure. For most of the MA star system’s measures, CMS retrospectively determines yearly star ratings based on the relative performance of all contracts over a past performance period (e.g., 2020 star ratings were determined using data that plans reported in June 2019 for the 2018 performance period). CMS uses a clustering algorithm—a method of grouping like-performing contracts—to identify “cut points” for assigning contracts to the five possible star levels for each of the measures (essentially forcing a five-group distribution). The weighted average of up to 45 individual-measure star ratings determined in this way constitutes a contract’s overall average rating (which, if at or above 3.75, will result in the bonus-level overall average rating of 4 stars or better).¹⁰ The retrospective clustering

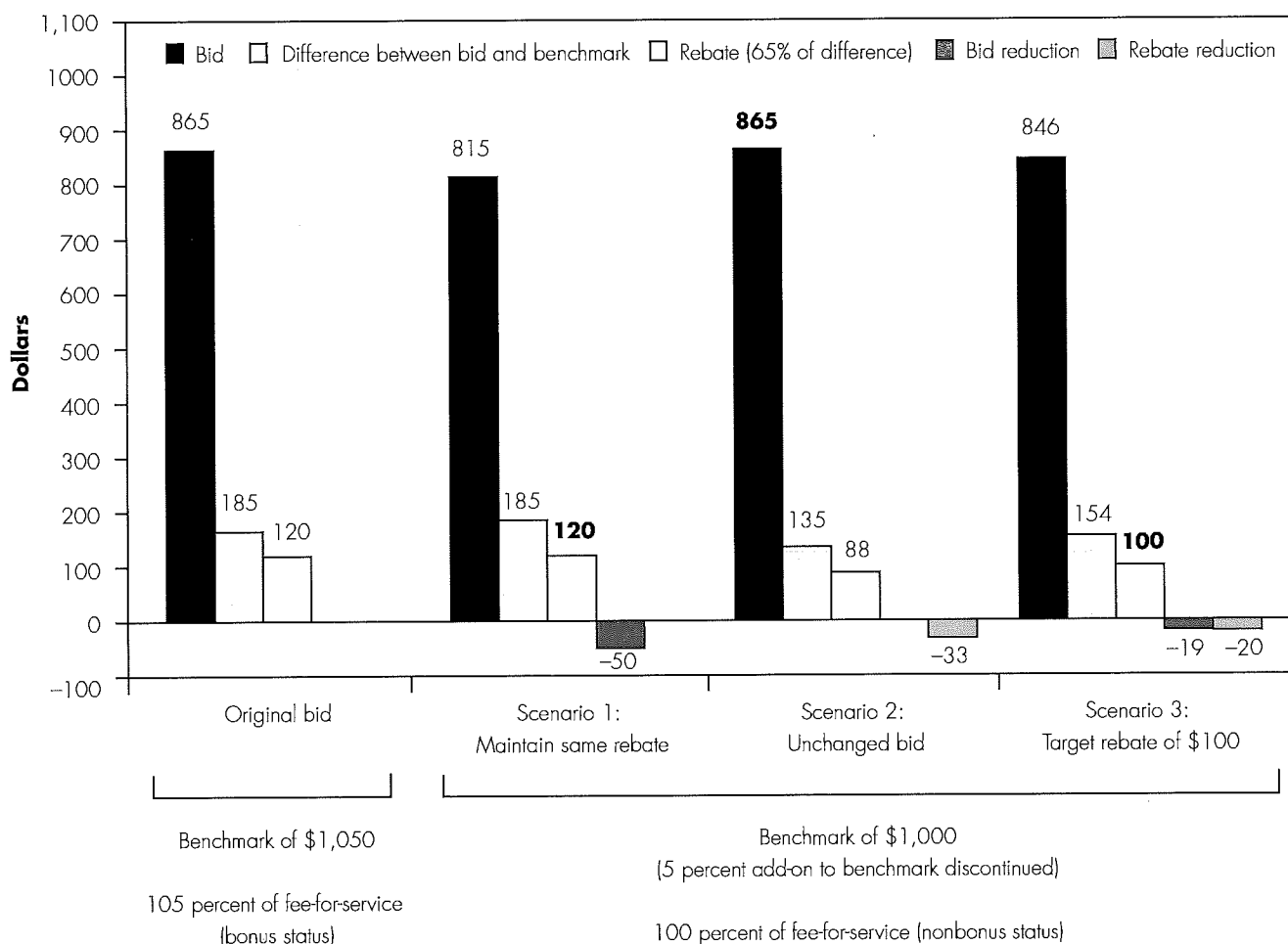
method decreases a plan’s ability to predict what star level will be assigned to a particular measure result in each year because, for example, the cut point separating a 3-star rating from a 4-star rating can be very different each year. A plan might have achieved a 4-star rating for a measure in one year that in the following year falls in the 3-star cluster, or a plan that had no change in results (or had a decline in performance) may move from a 3-star cluster to a 4-star cluster solely because of the distribution of results in the measurement year.¹¹ This unpredictability makes it difficult for providers and plans to manage their quality improvement efforts.

Unlike the current QBP, the MA–VIP is designed to reward or penalize a plan using a continuous, prospectively set scale for each measure. The performance scale could be set for each measure using different methods. For example, the performance-to-points scale can be set based on a broad distribution of historical data so that most entities have the opportunity to earn credit for their performance. Medicare can assess the performance-to-points scale annually and, if needed, revise the scale depending on whether expectations for quality achievement are met.¹² By making this scale continuous—that is, there are no cut points that need to be crossed in order for changes in quality to register—every improvement in quality is recognized by the MA–VIP. Unlike in the all-or-nothing QBP point system, in which a plan might determine that it is unable to achieve a 4-star (bonus) rating and the plan lessens its emphasis on quality improvement, in the MA–VIP, MA plans are always better off improving quality than not because the continuous scale provides incentives to achieve as high a score as possible for each measure.¹³

Prospectively set performance targets can drive quality improvement because plans are able to see how they will be rewarded for improvements in performance on measures. Under the MA–VIP, plans would be able to predict approximate rewards, given advance knowledge of the national performance-to-points scale for each measure (i.e., how their performance on measures translates to more points) as well as the approximate payment multiplier (i.e., the conversion of points to payment adjustments) for each peer group. The MA–VIP distributes rewards and penalties within a market area, and it would be administratively complex for CMS to accurately estimate and release these prospective payment multipliers (e.g., potentially 500 market areas with at least 2 peer groups in each). However, a couple of years after

**FIGURE
3-3**

**Illustrative example of how a plan could adjust bids and rebates if
5 percent add-on to benchmarks were to be discontinued**



Note: Illustrative example only, which assumes a plan in bonus status with rebate level of 65 percent, bidding for a population with a 1.0 average risk score, in a geographic area with a benchmark of 100 percent of fee-for-service spending, no cap on the benchmark level and in a non-double-bonus county. Totals may not sum due to rounding.

MA-VIP implementation, plans will have a general sense of how much of a reward they can receive for improved performance.

Distribute plan-financed rewards and penalties at a local market level

The MA-VIP is designed as a system of rewards and penalties. In this section, we discuss why and how the program will be financed through a portion of plan payments and the mechanism to fund a pool of dollars to distribute as rewards and penalties.

The MA-VIP will not be financed with added program dollars

When the Commission recommended a value incentive program for Medicare health plans in 2004, it was in the form of a system in which a small share of plan payments would be used to fund a pool of dollars that would redistribute money among plans based on their relative performance on quality metrics. No program dollars would have been added to fund the quality incentive program—unlike the current MA QBP, which uses additional program dollars to fund bonus payments.

An approach consistent with the Commission's long-standing recommendation in this regard achieves greater parity between MA and FFS (including ACO) quality incentive programs. This approach also results in savings to the Medicare program—reducing Part A expenditures and preserving trust fund dollars and providing savings to taxpayers, beneficiaries, and state Medicaid programs through reduced Part B expenditures and the premiums that all beneficiaries are obligated to pay to finance Part B.

The MA–VIP would be financed in the manner originally contemplated by the Commission: Quality incentive payments would be financed through a pool funded by a share of plan payments (as is currently done for the Medicare–Medicaid financial alignment demonstration plans through a withhold of up to 5 percent of total payments). The redesigned system would be a means of imposing financial pressure on health plans to increase their efficiency.

What is the potential effect of moving from a rewards-only to a plan-financed reward or penalty program?

For plans currently benefiting from higher benchmarks because they are in bonus status, the impact of discontinuing the use of added program dollars will depend on plans' bidding behavior and how they fare financially in the MA–VIP. Reduced Medicare revenues can affect plans' administrative expenses and profits, the level of extra benefits for enrollees, or payments to the plan's providers—or a combination of these factors.

Figure 3-3 illustrates various scenarios showing the effect on bids and rebates for a geographic area in which the benchmark changes from 105 percent (bonus) to 100 percent of FFS (nonbonus). A plan can decide (1) to change the plan bid to maintain the current rebate level; (2) leave the bid unchanged, with a resulting reduction in the value of extra benefits; or (3) modify the plan bid to achieve a certain target rebate level (\$100 in our illustrative example). (Though Figure 3-3 uses illustrative numbers, the rebate dollars as well as the amounts for bids and benchmarks are close to actual averages across MA.)

In the illustrative example, if the benchmark is \$1,000 rather than \$1,050, a company could decide to keep the rebate at \$120, which would require the plan to reduce its bid by \$50 (that is, a reduction in the cost of providing care, administrative costs, or profit, or a combination thereof) or it could decide to maintain a bid of \$865, which would reduce rebates by \$33 per month. Alternatively, the company could pursue a mixed strategy of only partly

reducing its rebates and could reduce its bid by less than \$50. If the company decided to set a rebate level of \$100, it would result in \$20 less in rebates for enrollees, but the company's bid would have to decline by only \$19.

The illustrative example of Figure 3-3 does not exactly convey what happens between one year and the next; it is more a comparison of bonus plans versus nonbonus plans in a given year in a given area. That is because, year over year, MA benchmarks increase due to inflation and other cost increase factors in MA and FFS (because FFS rates determine MA rates), or the benchmark in an area could change because of the change in the FFS quartile a county is assigned. If, for example, benchmarks were to rise by 5 percent year over year and QBP bonuses were no longer available, a plan could continue a rebate of \$120 without any change to its bid. Beginning in 2021, plans will no longer be required to pay the 2 percent health insurer fee instituted by the Affordable Care Act of 2010. The fee is 2 percent of revenue (the CMS payment of the plan's bid plus the rebate dollars). In the illustrative example of Scenario 3, 2 percent of revenue would average about \$19 per member per month (2 percent of revenue of \$946 (\$846 + \$100)), meaning that in the last example, the \$19 bid reduction to arrive at a rebate of \$100 could be offset entirely by the added revenue resulting from the repeal of the health insurer fee. However, bids also change from year to year for various reasons—such as a plan's practice of passing on benchmark increases to its providers, a change in the provider network to include higher cost providers, or changes in the competitive environment that would put pressure on a plan to increase its extra benefits.

Our past analysis of actual bidding behavior suggests that plans have a strong motivation to try to avoid reductions in extra benefits while at the same time not necessarily increasing extra benefits when revenue from CMS increases. That is, declines in a plan's revenue do not result in a dollar-for-dollar decline in extra benefits, nor do increases in revenue result in dollar-for-dollar increases in extra benefits. Our previously presented analysis of the bids for 2019 shows that most of the extra dollars from bonus payments were not used to provide extra benefits to MA enrollees, and only those plans that saw a decline in their benchmarks due to the loss of bonus status reduced their costs of providing the basic Medicare benefit package (see Figure 3-4, p. 64). The text box (pp. 64–67) provides additional details about the actions plans took between 2018 and 2019 in reaction to changes in MA revenue.

Changes in bids between 2018 and 2019 show that plans reduce administrative costs and profits to maintain extra benefits

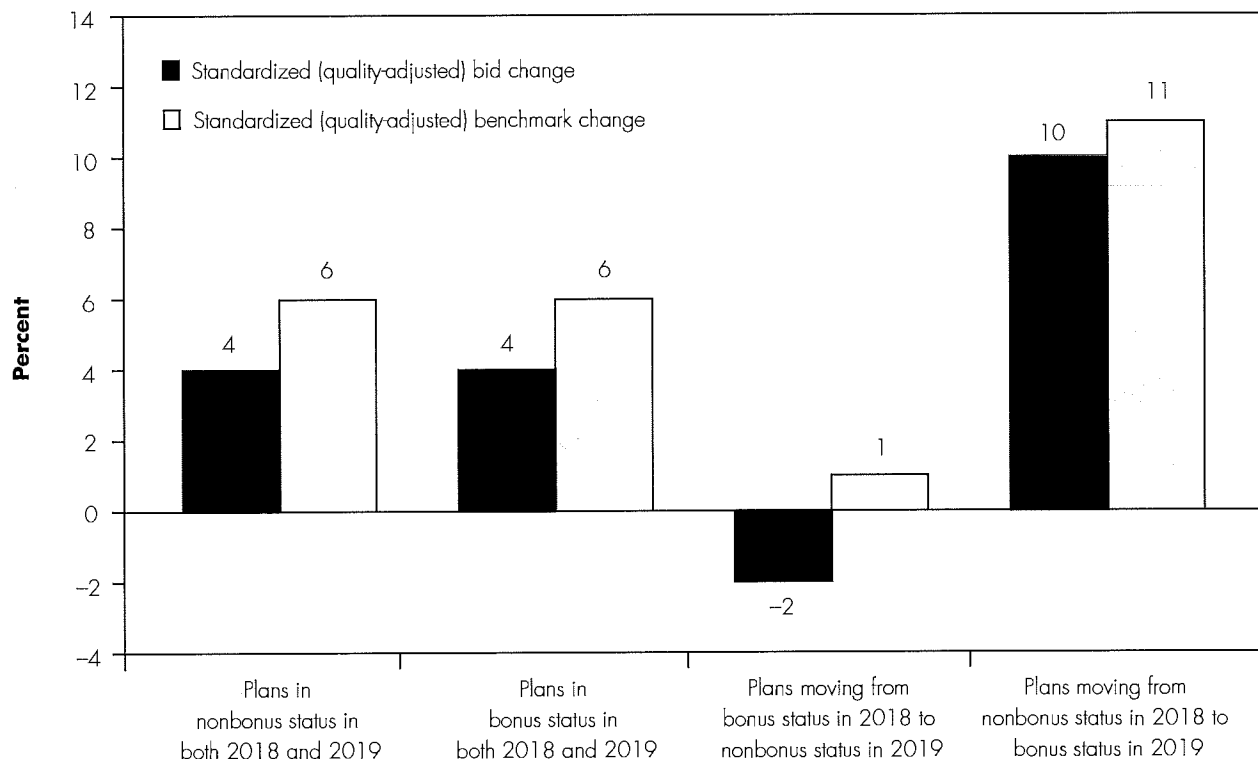
Figure 3-4 shows the change in bids and benchmarks between 2018 and 2019 based on plans' bonus status or change in bonus status. The bids and benchmarks are standardized amounts, representing amounts for a population of average risk. The "standardized bid change" amounts show the level of plans' medical inflation for the Medicare Part A and Part B benefit package (the cost of the benefit, administration, and profit). For plans that maintained the same bonus (or nonbonus) status

between 2018 and 2019, the cost of providing the Medicare benefit—including administrative costs and profit—rose by a risk-standardized 4 percent. For such plans, benchmarks increased 6 percent (for a population of average risk). For plans that had an increase in their Medicare payments because they moved from nonbonus status to bonus status, the reported cost of providing the Medicare benefit rose by 10 percent—over twice the increase for the other bonus status categories of plans shown in Figure 3-4. The

(continued next page)

FIGURE 3-4

Additional quality bonus payments resulted in higher bids, while plans losing bonus status reduced their bids for the Medicare benefit package between 2018 and 2019



Note: Special needs plans are excluded. Excludes plans with changes in segments (subplan classifications) that materially differ between the two years. All bid data pertain to the Medicare Part A and Part B benefit package.

Source: MedPAC analysis of 2019 Medicare Advantage bid data.

Changes in bids between 2018 and 2019 show that plans reduce administrative costs and profits to maintain extra benefits (cont.)

rise in medical inflation for these plans (10 percent) nearly matched the rise in quality-adjusted benchmark levels (11 percent). In contrast, plans moving from bonus status to nonbonus status reduced their cost of providing the Medicare benefit in the face of only a small increase in the benchmark.

Table 3-3 breaks down the components of the payment changes for plans' bonus status categories, showing plans that lost bonus status did not reduce their level of extra benefits but changed other factors in their bids. The table compares (1) actual bids (not standardized for risk—i.e., representing the actual costs plans expect to incur, based on the expected risk of the plan's enrollees) against (2) benchmarks that have been risk adjusted using the plan's projection of the risk of its enrollees. The value of rebates offered when a plan

bids below the benchmark is established by comparing risk-adjusted amounts because Medicare's payments to a plan are risk adjusted (i.e., the plan's risk-adjusted payment is more or less than the Medicare base payment). The difference between the expected payment from Medicare and the expected cost of providing the benefit is the basis for determining the rebate amount.

Table 3-3 shows that, in the case of plans leaving bonus status (bonus to nonbonus), their benchmarks increased (reflecting a base benchmark increase of 1 percent) and the projected risk scores increased for these plans (risk score data not shown in table). Such plans had an enrollment-weighted benchmark increase of \$46, of which \$32 (or 70 percent) was allotted to the rebate computation, producing a monthly beneficiary

(continued next page)

**TABLE
3-3**

In 2019, for plans newly receiving bonus payments, only a small share of the benchmark financed extra benefits, but plans leaving bonus status applied most of their benchmark increase to extra benefits

	Bonus status unchanged from prior year	Plans moving from bonus to nonbonus status	Plans moving from nonbonus to bonus status
Risk-adjusted benchmark increase	\$72	\$46	\$108
Risk-adjusted bid increase	\$48	\$14	\$83
Marginal addition to rebate computation (benchmark minus the bid)	\$24	\$32	\$26
Value of extra benefits to beneficiaries (50 percent to 70 percent of rebate, based on stars)	\$16	\$21	\$17
Share of benchmark increase applied to rebates	33%	70%	24%
Components of the risk-adjusted bid increase			
Dollar change in net medical expenses	53	\$30	\$59
Dollar change in administrative costs	-\$10	-\$5	-\$10
Dollar change in Medicare margin	\$4	-\$10	\$33

Note: Special needs plans are excluded. Table excludes plans with changes in segments (subplan classifications) that materially differ between the two years. All bid data pertain to the Medicare Part A and Part B benefit package. Dollar figures are per member per month amounts. Components may not sum to totals due to rounding.

Source: MedPAC analysis of 2019 Medicare Advantage bid data.

Changes in bids between 2018 and 2019 show that plans reduce administrative costs and profits to maintain extra benefits (cont.)

rebate amount of \$21. These plans' bids increased very little (by \$14); they reduced their margins by an average of \$10 per member per month; they reduced their administrative costs; and their Medicare Part A and Part B medical expenses increased less than those of other plans (by \$30). For the two other plan categories, plans remaining in the same bonus status and plans changing from nonbonus to bonus status, a third (\$24 of \$72) or less (\$26 of \$108) of the benchmark increase was applied toward the rebate computation, respectively. In the nonbonus-to-bonus category, 30 percent of the increased benchmarks (\$33) was used to increase plan margins, and payments for Medicare-covered health care services increased. Of the three components of the bid for the Medicare Part A and Part B benefit—medical costs, administrative costs, and margin (profit)—the administrative cost component decreased for all categories shown in the table. (Because of the increase in margins, it may have been necessary to reduce administrative costs to maintain a medical loss ratio—

the share of the bid going toward medical costs—at 85 percent or less, as required of Medicare Advantage (MA) plans.) Other factors may also play a part in the differences we see in comparing 2018 bids with 2019 bids—including the type of plans involved (preferred provider organizations tend to have higher bids than HMOs) or the geographic area involved (areas with high fee-for-service (FFS) utilization tend to have lower bids in relation to area FFS levels, allowing plans to offer richer benefits). Thus, the effect of bonus funding on plans' bids in the 2018 to 2019 period may not be the same in a different period.

The illustrative example in Figure 3-3 (p. 62) shows possible plan behavior in the face of reduced Medicare revenue and the potential effect on plan bids and extra benefits. Figure 3-4 (p. 64) and Table 3-3 (p. 65) show actual plan behavior in 2019 in the face of declines or increases in revenue. The actual behavior suggests that plans will tend to maintain a similar level of extra benefits from one year to the next and will forgo profits

(continued next page)

Mechanism to fund pool of dollars to distribute as rewards and penalties

In the MA–VIP design, the mechanism to fund the pool of dollars—through which rewards and penalties would be distributed—could be structured in (at least) two ways: through a withhold of plan payments that is returned in a lump sum determined on the basis of quality performance or through a payment adjustment that would increase or decrease all plan payments by a certain percentage based on their quality performance.

If the MA–VIP were funded through a withhold, plan payments would be reduced by 2 percent, for example, for the year in which plan performance is assessed. We assume data collection would end six months after the end of the performance year, including encounter data collected through that point, and plan performance would

be converted to an MA–VIP payment amount. MA–VIP payments would redistribute the 2 percent withhold funding based on quality scores and could be sent to plans in a lump sum based on quality performance. MA–VIP payments that are smaller than a plan's 2 percent withhold would effectively be a penalty, while payments that are larger would effectively be a reward.

An alternative to the withhold approach is to use a payment adjustment mechanism, as is done in the hospital value incentive program (HVIP). Our illustrative MA–VIP model uses terms consistent with applying a payment adjustment to future plan payments. These payment adjustments would be set equal to 2 percent of plan payments, but there would be no withholding of plan payments. Instead, plan quality would be assessed during the performance year, data collection would be completed

Changes in bids between 2018 and 2019 show that plans reduce administrative costs and profits to maintain extra benefits (cont.)

if necessary but use only a portion of added revenue to finance extra benefits. Figure 3-4 (p. 64) and Table 3-3 (p. 65) show that for plans newly in bonus status (moving from nonbonus to bonus), bonus payments are not entirely used to provide extra benefits to plan enrollees. Instead, additional dollars in 2019 were used to increase margins and payments to providers. Plans whose bonus status did not change used a greater share of payment increases to apply to the rebate computation and less toward increasing their margin; the greatest share of the payment increases went toward provider payments and other components of the cost of providing the Medicare benefit (such as quality improvement activities). For plans losing their bonus status (but still receiving higher payments because of higher risk scores), the largest share (70 percent) of their increased payment was applied to maintain or improve their level of extra benefits through the rebate. In the face of financial pressure, such plans reduced their margins and reduced their cost of providing the Medicare benefit.

The current Medicare plan payment rates finance a generous level of extra benefits for enrollees, which averages \$122 per enrollee per month in 2020. We expect changing from the quality bonus program (QBP) financing method of added program dollars to an MA value incentive program financed without added program dollars would result in a relatively small decline in the record-level rebates for MA enrollees. If the added program dollars of the QBP had been discontinued in 2020 (and assuming plans made no adjustments to their bids), we computed the potential decline in rebates to be \$27—similar to the \$33 amount in the illustrative example (Figure 3-3, p. 62).¹⁴ We estimate that, stated in relation to the current level of extra benefits, if there had been a reduction of \$6 billion in available dollars, the plan behavior described in Table 3-3 (p. 65) would have resulted in a reduction in extra benefits in the range of \$6 to \$17 per member per month. For 2020, then, the average level of extra benefits would have declined from \$122 to a range of \$105 to \$116 per month—similar to, or somewhat higher than, the \$107 level of extra benefits in the preceding year, 2019. ■

and plan performance would be converted to an MA–VIP payment adjustment the next year to apply to payments the following year. The MA–VIP payment adjustments would be the net of a 2 percent funding pool, generating negative payment adjustments (penalties) and positive payment adjustments (rewards). Under this scenario, plan performance would be assessed in year 1. After a one-year lag to collect data and calculate the size of the payment adjustments (year 2), adjustments would be applied to monthly plan payments in year 3.

Distribute rewards and penalties within local market areas

With MA plan quality evaluated at the local market level, it would be possible to distribute rewards and penalties to plans either within each market or nationally. The remainder of this section reflects the Commission's

consideration of local and national approaches to distributing rewards and penalties, concluding with the Commission's support for distributing rewards and penalties within each local market area.

Distributing rewards and penalties within each market area means that the value of rewards equals the value of penalties in each market, and net MA–VIP payments are zero in every market area. Under this approach, for each peer group, the parent organization with the highest quality score in the market receives the greatest reward, and the organization with the lowest score in the market receives the greatest penalty. Distributing rewards and penalties this way provides an incentive for each parent organization to improve quality within the market and for each peer group in that market. Thus, plans are rewarded for their performance in each market.

**TABLE
3-4****Comparison of favored local market distribution of MA-VIP rewards and penalties versus national distribution**

Considerations	Local market distribution	National distribution
Quality improvement incentives	<i>Improve quality in every market</i>	<i>Improve quality in every market, but plan offerings may be more numerous in markets with higher average MA quality</i>
Geographic neutrality	<i>Geographically neutral: Does not favor MA plan participation in some markets over others</i>	<i>Not geographically neutral: Favors MA participation in markets with higher average quality</i>
Neutrality with FFS Medicare	<i>Maintains neutrality: Neither MA nor FFS Medicare is favored in any market</i>	<i>Does not maintain neutrality: Favors MA in high MA quality markets and favors FFS in low MA quality markets, both without regard to FFS quality in those markets</i>
Alignment with beneficiary plan options	<i>Aligned with beneficiary plan options: Best-performing plans receive rewards and worst-performing plans receive penalties in each area</i>	<i>Partially aligned with beneficiary plan options: Performance in relation to a national standard means that worst-performing plans in an area may receive rewards and best-performing plans may receive penalties</i>
Plan accountability for exogenous market conditions	<i>Plans are not accountable for market conditions outside their control: Plans are not held accountable for exogenous market conditions</i>	<i>Plans are accountable for market conditions outside their control: Plans are held accountable for exogenous market conditions</i>
Alignment of rewards with local or national quality performance	<i>Rewards aligned with local market performance: Quality scores reflect plans' effectiveness at improving quality in each market</i>	<i>Rewards aligned with national performance: Quality scores reflect plans' effectiveness at improving quality in each market but also the underlying market conditions that affect average MA quality nationally</i>
Plan administrative burden to track performance	<i>Low performance-tracking burden: Plans track only a few MA competitors within their local market to assess relative performance and calibrate quality goals</i>	<i>High performance-tracking burden: Plans track all MA competitors in the country to assess relative performance and calibrate quality goals</i>

Note: MA-VIP (Medicare Advantage value incentive program), FFS (fee-for-service).

Distributing rewards and penalties at the market-area level holds constant the market conditions that are outside of a plan's control (e.g., availability of safety net programs like Medicaid and food assistance, transportation infrastructure, the level of social risk factors in the population, and the underlying organization of providers in each market). National distribution would hold plans accountable both for their performance and for local market conditions. In addition, national distribution could result in rewards for all plans in some markets and penalties for all plans in other markets because payments would be redistributed from markets with lower MA quality to markets with higher MA quality. Over time, increased payments in markets with higher average

MA quality and decreased payments in markets with lower average MA quality could skew the geographic distribution of plan offerings.

The Commission's HVIP distributes rewards and penalties nationally, meaning a pool of dollars is distributed to hospitals based on their quality performance, regardless of the hospital's location. Under this approach, rewards and penalties may not be distributed evenly across the country. In contrast to hospitals, MA plan sponsors can change the markets in which they operate each year. Because of this flexibility and certain benefits described below (e.g., not holding plans accountable for exogenous market conditions, not favoring MA or FFS in any market,

better aligning with beneficiary plan options, and lower administrative burden for plans tracking performance), we designed the MA–VIP to distribute rewards and penalties by market. Table 3-4 provides a comparison of the differences between local market–level and national distribution of rewards and penalties. The Commission also considered (but did not recommend) a blended market level–national approach that would enable a share of the rewards to be distributed to the highest performing plans in the market from a local reward pool financed by a portion of all plan payments in the market area. Under the blended approach, the remaining share of rewards would be distributed from a national pool of dollars financed by a portion of all plans’ payments across the country.

The Commission has maintained a standard of not favoring either the MA program or FFS Medicare with respect to their payment systems or monitoring and compliance activities. Ideally, we would compare MA plan quality with local FFS quality in each market and reward MA plans that provide higher quality than FFS in the area. However, such a comparison between MA and FFS is currently not feasible.¹⁵ Distributing MA–VIP rewards and penalties by market does not favor either the MA program or FFS Medicare because all MA–VIP plan rewards and penalties are confined within each market, having a zero-dollar net effect in every market. In contrast, national distribution of rewards and penalties favors the MA program in markets with high average MA performance and favors FFS Medicare in markets with low average MA performance, regardless of whether the MA performance is better than local FFS performance. Some or all MA plans in markets with low average quality may offer higher quality than local FFS Medicare, yet those plans would receive a penalty under national distribution. Conversely, MA plans performing below local FFS in markets with high average quality would receive a reward under national distribution. Until FFS comparisons are possible, distributing rewards and penalties within each market maintains neutrality between the two programs.

Medicare beneficiaries generally do not move their residence to a different market on the basis of their local Medicare FFS and plan options. Distributing rewards and penalties by market aligns MA–VIP payments with the best MA plan options in each market, providing a payment increase to the best MA performers and payment decrease to the worst performers. Distributing rewards and penalties nationally maintains MA plan performance in each market as the basis of evaluation, but could provide rewards to

the worst performing plan options or penalties to the best performing plan options available to beneficiaries in a given market.

MA quality scores are a function of factors under the plan’s control (e.g., provider network management and incentive programs) and market conditions outside an MA plan’s ability to control (such as the availability of safety net programs like Medicaid and food assistance). Factors within a plan’s control can differentiate plan quality scores within a market, but market conditions outside a plan’s control tend to explain why average MA quality varies across markets, including differences in average MA quality in markets with the same set of parent organizations. Our initial modeling shows wide variation in average MA quality across markets (see Figure 3-5, p. 80, and Figure 3-6, p. 81). Distributing rewards and penalties within each market would not hold plans accountable for market conditions that are outside of their control, and differences between parent organization quality scores within each market would generally reflect plans’ effectiveness in improving quality in that market. Distributing rewards and penalties nationally would hold plans accountable for factors outside their control, and differences in quality scores would jointly reflect differences in market conditions and differences in plans’ effectiveness in improving quality in a given market.

Aside from the question of accounting for differences in market factors, a further consideration is whether rewards and penalties should be tied to a plan’s local performance or the plan’s performance in relation to a national standard. This consideration does not involve whether performance results should be reported locally or nationally, but whether local or national performance is more justified as a basis for distributing rewards and penalties. For example, if rewards and penalties were distributed locally, some parent organizations with quality scores above the national average would receive a penalty for performing below average in their market, while some parent organizations with quality scores below the national average would receive a reward for performing above average in their market. The level at which rewards and penalties are distributed determines whether plans are held accountable for local market conditions outside of their control. If plans are held harmless for exogenous factors that exist in their local markets, as with local distribution of rewards and penalties, the perceived misalignment of rewards and penalties across markets is not a concern (plan performance is assessed in comparison to local

plans operating under the same market conditions). Conversely, if plans are judged on all factors affecting their quality score including market conditions outside their control, as with national distribution, then rewards and penalties would be aligned across markets, but in such a case, the use of national standards for determining rewards and penalties should be weighed against the other considerations noted in Table 3-4 (p. 68).

A final consideration is the ability of plans to track their performance and assess any MA–VIP rewards or penalties. Plans attempting to track and set goals for MA–VIP rewards or penalties will track not only their own performance but also the performance of competitors. When MA–VIP rewards and penalties are distributed by market, the burden of performance tracking is relatively low because competition is limited to the other plans in the same market; however, this burden increases when distributing MA–VIP rewards and penalties nationally, as competition encompasses every plan in the country. Applying a local and national blend to distribute rewards and penalties would impose the highest burden, requiring plans to track their local performance and national performance.

The Commission assessed the merits of using a blended approach in which a share of a plan’s ultimate reward or penalty would be based on a nationally distributed pool of dollars, while the rest would be based on a pool of dollars distributed within the plan’s local market. The relative size of the national pool and market pools of dollars would be determined by weighting the local and national components of the blend. The overall reward or penalty for a parent organization in a given market would depend on the weight of each component as well as the relative magnitude of rewards and penalties garnered by the plan from the national pool and from the local market pool. Under a blended approach, a plan with above-average quality performance in its market but below-average quality performance nationally would receive a reward for its high market-area quality performance and a penalty for its low national quality performance.

The Commission generally does not support a blended approach, which would share the attributes of both a local and national approach: It would not maintain geographic neutrality or neutrality between MA and FFS programs (features of a pure local approach); it would not align rewards and penalties with beneficiary local plan options (local performance) or national performance; it would hold plans partially accountable for market conditions outside

their control; and it would require the greatest complexity for plans to assess their performance, by requiring them to assess their local performance and their national performance. Given these considerations, the Commission supports the distribution of rewards and penalties within each local market over a national approach.

Illustrative scoring and payment adjustments under the MA–VIP model

To analyze potential MA plan performance under the MA–VIP design, we modeled scoring and calculating payment adjustments in a subset of market areas based on currently available data. To account for differences in the social risk factors of plan populations, within each market area, we stratified each parent organization’s enrolled population into two peer groups: fully dual-eligible enrollees and all other enrollees. We converted the performance of each MA plan peer group to an MA–VIP payment adjustment that converts to a reward or penalty. (See text box on converting quality performance to rewards or penalties, pp. 72–73.) For many market areas, we do not have sufficient data that would allow us to calculate MA plan performance on the full set of our MA–VIP model measures; however, our model results show that the MA–VIP design elements can feasibly be incorporated into a redesigned and improved quality incentive program to replace the QBP. Also, as intended in the design of the MA–VIP methodology, the peer group with more social risk factors receives a relatively higher reward for higher quality. The modeling results also demonstrate that, as compared with the QBP, the MA–VIP reduces the disparity between fully dual-eligible enrollees and other populations when determining how financial incentives are distributed.

Calculate plan performance on a small set of measures

We modeled the MA–VIP using MA plan performance on 6 of the 12 measures presented in the illustrative measure set in Table 3-2 (p. 58). We were limited in the data available to calculate meaningful (reliable) measure results for the MA–VIP reporting unit (parent organization in a market area) because MA quality measurement is currently done at the contract level and some contracts span multiple market areas.

The lack of complete MA encounter data also limits the administrative data–based measures we can include in

the MA–VIP model. The Commission has previously recommended that, given the value of complete encounter data, CMS should improve plan performance metrics to include assessments of data completeness, implement a payment withhold to introduce the financial incentive to submit complete and accurate data, and require submissions of providers’ claims directly to Medicare administrative contractors if encounter data performance thresholds are not met (Medicare Payment Advisory Commission 2019b). Through its consideration of the recommendation, the Commission expressed broad support for using the encounter data in many applications to improve incentives for increasing the completeness and accuracy of the data.

For ACS hospitalizations, we can supplement inpatient encounter data with MA inpatient data reported in the Medicare Provider Analysis and Review (MedPAR) file.¹⁶ However, for ACS ED visits, there is no other data source to supplement outpatient encounter data, so we determined that we cannot measure ACS emergency department visits at this time.¹⁷ We did not include the readmissions measure in the model because of technical issues converting the encounter data that would be used to calculate risk-adjusted readmissions results.

For the six measures that we can include in our MA–VIP model, we calculated measure results for each reporting unit (parent organization in a market area and, where relevant, peer group) using four available data sources: (1) encounter data MA plans submit to CMS supplemented with other administrative data sources (i.e., MedPAR hospital inpatient data reported by hospitals on all Medicare FFS and MA inpatient stays); (2) beneficiary-level patient-reported outcomes data from the HOS (collected by certified survey vendors on behalf of plans); (3) beneficiary-level patient/enrollee experience data from CAHPS surveys (collected by certified survey vendors on behalf of plans); and (4) beneficiary-level data on HEDIS measures that plans submit to CMS. The measure calculations are based on existing Commission, CMS, or HEDIS measure specifications. We also applied existing CMS or industry minimum sample sizes to determine whether a reporting unit had complete performance results. Table 3-6 (p. 74) summarizes the measure calculations used in our MA–VIP model.

To increase the number of observations in our model, we pooled three years of data (2015 to 2017) for most of our measure calculations. This amount of data was especially

important to increase the number of reporting units that would meet the minimum sample sizes for surveys to be scored in MA–VIP modeling since we use beneficiary-level survey responses that are based on a sample of enrollees at the contract level and rescore them into results for the MA–VIP reporting units based on where the enrollee resides. (Under the MA–VIP, each parent organization meeting a minimum enrollment threshold in a market area would be required to work with a third-party survey vendor to collect CAHPS and HOS responses from enrollees at the market-area level, as opposed to the contract level.)

In implementing the MA–VIP, policymakers will need to determine how many years of data to use in measure calculations. Using the most recent year of data holds MA plans accountable for the quality of their most recent care provided to enrollees and is likely a better predictor of the quality of care in the subsequent year. Using measure results based on multiple years of data reduces random variation from smaller sample sizes and allows Medicare to measure the quality of care for low-volume plans. However, to reward performance that improved (or declined) over the multiple-year period, the model could weight recent-year performance more heavily than performance in earlier years. The model could also use the most recent year of data for plans that meet minimum sample size requirements and multiple years for those that do not meet the minimum sample size in the most recent year. One disadvantage of this approach is that small plans would be held accountable for their performance through multiple years, while large plans would be held accountable for only one year of performance, which could be perceived as applying different accountability standards to small versus large plans.

The key components of our model calculate performance within a local market area with stratified scoring and separate pools of dollars for fully dual-eligible enrollees and all other enrollees. Thus, we calculate separate measure results for a reporting unit’s fully dual-eligible population and all-others population. Consistent with the Commission’s principles for quality measurement, the specifications for the ACS hospitalization measure we developed do not include social risk factors (such as dual eligibility for Medicare and Medicaid) in the risk adjustment model. Therefore, we calculate ACS hospitalization results for both peer group populations of each reporting unit that meets the minimum sample size requirement (i.e., 150 fully dual-eligible enrollees and 150

Using peer groups to convert quality performance to rewards or penalties in a local market area

In the following example, a local market area has three Medicare Advantage (MA) parent organizations (referred to in this example as “three MA plans”) for which to calculate performance measure results. We stratify each plan’s enrollee population into two peer groups: fully dual-eligible enrollees (Peer Group 1) and all other enrollees (Peer Group 2). Following several steps, we convert each of the MA plans’ peer group quality measure performance to a payment adjustment and combine the peer groups’ payment adjustments into one total Medicare Advantage value incentive program (MA–VIP) adjustment. Specifically, we followed six steps:

Step 1: For each peer group, calculate each MA plan’s performance on the quality measures; this step produces a performance rate for each plan’s peer groups for each measure. The calculations are based on either beneficiary-level administrative data or survey data.

Step 2: Convert each MA plan’s performance on the quality measures for each peer group to points based on the same continuous performance-to-points scale (nationally determined).

Step 3: Calculate the weighted average of each MA plan’s points on the quality measures to determine total MA–VIP points for each peer group. (Assume higher weighting for outcome measures.)

Step 4: For each peer group, create a pool of expected MA–VIP payments to plans, based on a specified percentage tied to plan payments for each peer group (e.g., 2 percent of each plan’s payments for their peer group’s population).

Step 5: For each peer group, calculate the payment multiplier or percentage adjustment to payment per MA–VIP point, which converts total MA–VIP points to dollars and results in spending each group’s pool of dollars defined in Step 4.

Payment multiplier = MA–VIP pool for peer group
/ sum of (each MA plan’s payment tied to the peer

group × each MA plan’s total MA–VIP points for the peer group)

Step 6: Compute each MA plan’s adjustment for the coming year based on past performance and its peer groups’ payment multiplier.

MA plan’s total MA–VIP adjustment = (Peer Group 1 payment multiplier × MA plan’s total MA–VIP points for Peer Group 1) + (Peer Group 2 payment multiplier × MA plan’s total MA–VIP points for Peer Group 2)

Table 3-5 illustrates the conversion of MA–VIP points to payment adjustments using peer grouping in a local market area with three MA plans that have different numbers of fully dual-eligible and other enrollees. We calculate quality measure results based on administrative and survey data for each plan’s fully dual-eligible enrollees (Peer Group 1) and all other enrollees (Peer Group 2) for each of the five measure domains. Using the same nationally determined continuous performance-to-points scales, we convert each peer group’s quality performance to points for each domain. We average each plan’s performance by peer group to determine MA–VIP total points for each plan’s peer groups. The table shows that MA Plan A earns the highest performance across both peer groups (8 points). MA Plans B and C both earn lower points for their fully dual-eligible population (4 points) compared with their other-enrollee population (6 points).

We create a pool of dollars based on 2 percent of each of the MA plan’s payments tied to each of the peer groups. Since MA Plan C has the largest number of enrollees, its contribution to the pool of dollars is largest. The pool to be redistributed for Peer Group 2 (other enrollees) is larger than Peer Group 1’s pool because more enrollees and payments are in Peer Group 2. For each peer group, we calculate a payment multiplier or percentage adjustment to payment per MA–VIP point. The payment multiplier for each peer group is the group’s pool of dollars divided by the

(continued next page)

Using peer groups to convert quality performance to rewards or penalties in a local market area (cont.)

sum of each plan's total payments times their MA–VIP total points. Because Peer Group 1 has a larger point multiplier than Peer Group 2, the plan with higher performance for its fully dual-eligible enrolled population can earn a higher reward.

We calculate payment adjustments based on each peer group's MA–VIP points and payment multiplier. In total, MA Plan A has the highest performance for both

peer groups and so earns a reward of 1.21 percent, net of its 2 percent of payment that went into the pool. On net, MA Plan A earns a reward of \$3.5 million for Peer Group 1 and a reward of \$1.3 million for Peer Group 2, for a total reward of \$4.8 million. MA Plans B and C both receive small penalties because they receive fewer points for both their fully dual-eligible enrollees and all other enrollee populations. The entire pool of dollars is distributed to the MA plans in the market. ■

**TABLE
3–5**

**Converting MA–VIP points to payment adjustments
in a local market area: An illustrative example**

	Peer Group 1 (fully dual-eligible beneficiaries)			Peer Group 2 (all others)		
	Plan A	Plan B	Plan C	Plan A	Plan B	Plan C
Number of beneficiaries	10,000	100,000	54,000	20,000	620,000	820,000
MA–VIP total points (Steps 1–3)	8	4	4	8	6	6
Plan payments tied to each peer group's beneficiaries	\$200M	\$2,000M	\$1,080M	\$200M	\$6,200M	\$8,200M
2 percent of plan payments tied to each peer group's population	\$4M	\$40M	\$21.6M	\$4M	\$124M	\$164M
Total pool of dollars for peer group (Step 4)		\$65.6M			\$292 M	
Payment multiplier for peer group [group's pool / sum (plan payments x points)] (Step 5)		0.47%			0.33%	
MA–VIP payment adjustments [points x multiplier] (Step 6)	3.77%	1.89%	1.89%	2.65%	1.99%	1.99%
MA–VIP payments [multiplier x plan payments]	\$7.5M	\$37.7M	\$20.4M	\$5.3M	\$123.4M	\$163.3M
Net payments	\$3.5M	–\$2.3M	–\$1.2M	\$1.3M	–\$0.6M	–\$0.7M
Total MA–VIP payment adjustment (net after 2 percent of payment)						
Plan A			+1.21% (+\$4.8M)			
Plan B			–0.03% (–\$2.9M)			
Plan C			–0.02% (–\$2.0M)			

Note: MA–VIP (Medicare Advantage value incentive program), M (million). This example assumes a local market area has three Medicare Advantage plans. Fully dual-eligible beneficiaries qualify for a full range of Medicaid benefits. MA–VIP total points range from 0 to 10 points. Totals may not sum to components due to rounding.

**TABLE
3-6****Illustrative MA-VIP model: Calculating performance on a small set of available measures**

Domain	ACS hospital use	Patient-reported outcomes	Patient/enrollee experience	Staying healthy and managing long-term conditions
Measures	ACS hospitalizations	Improved or maintained physical health status Improved or maintained mental health status	Getting needed care Rating of health plan	Breast cancer screening
Years / time period	2015–2017	One time period: change between 2015–2017	2015–2017	2015–2017
Minimum sample size	150 enrollees ^a	30 completed enrollee surveys ^b	100 completed enrollee surveys ^c	30 women meeting inclusion criteria ^d
Risk or case-mix adjustment	Risk-standardized rates based on method developed by RTI International for the Commission	CMS HOS case-mix adjustment	CMS CAHPS [®] case-mix adjustment	Not adjusted
Modeling data sources	Encounter data, MedPAR	Beneficiary-level HOS survey data	Beneficiary-level CAHPS survey data	Beneficiary-level HEDIS [®] data
Stratification	Fully dual eligible All others	None	None	Fully dual eligible All others

Note: MA-VIP (Medicare Advantage value incentive program), ACS (ambulatory care-sensitive), MedPAR (Medicare Provider Analysis and Review), HOS (Health Outcomes Survey), CAHPS[®] (Consumer Assessment of Healthcare Providers and Systems[®]), HEDIS[®] (Healthcare Effectiveness Data and Information Set[®]). "Minimum sample size" is the number of observations required across the years included or time period to be included in the MA-VIP model. The MA-VIP calculates stratified scoring for two groups (fully dual-eligible enrollees and all other enrollees). The MA-VIP scores the ACS hospitalization and breast cancer screening measures with separate rates for each of the groups, but the same rate across both groups for the two survey-based measure domains because case-mix adjustment factors address group differences.

^a Based on minimum sample size for similar HEDIS measure.

^b CMS statement regarding minimum sample to determine differences among plans.

^c RAND-determined minimum sample size for health plan CAHPS.

^d National Committee for Quality Assurance HEDIS measure-specific minimum (e.g., women ages 50–74).

non-fully dual-eligible enrollees for ACS hospitalizations). For the two survey-based measure domains (CAHPS and HOS), the CMS methodology includes eligibility for Medicaid in the case-mix adjustment. We therefore use the result based on the entire MA population to determine the score for both peer groups in the reporting unit. For example, a reporting unit's result of 65 percent on the HOS "improved or maintained physical health" measure based on a sample of all MA enrollees would apply to MA-VIP scoring in both groups. CMS has identified differences between breast cancer screening (BCS) rates between the two populations, so we also calculated a separate BCS rate for each group.

For the patient-reported outcome measures, we followed CMS's method of producing case-mix-adjusted HOS measure results to determine the share of enrollees showing maintenance or improvement of their physical health and mental health. (Enrollees were surveyed in 2015 and again in 2017 to determine changes in health status for those remaining in the same MA contract over the two-year period.) CMS currently collects survey responses from a sample of enrollees selected at the contract level, not at the parent organization and market-area levels. When one contract's service area consists of counties in two noncontiguous states, such as Hawaii and Iowa, the HOS results for that contract are based

on a sample of enrollees residing in Hawaii and Iowa. In contrast, under the MA–VIP model, we mapped individual enrollee HOS results, which include person-level identifiers, to a local market area using the enrollee’s county of residence and to a parent organization using the enrollee’s plan identifier. To be included in the MA–VIP model, a parent organization in a market area must have 30 or more beneficiaries with HOS results attributed to the organization. Based on the survey responses attributed to an MA–VIP reporting unit, we calculated case-mix-adjusted HOS measure results (e.g., improved or maintained physical and mental health status) for each reporting unit.

We used 2015 to 2017 beneficiary-level CAHPS survey responses to calculate case-mix-adjusted patient experience results for each MA–VIP reporting unit. As with the HOS, CMS currently collects survey responses from a sample of enrollees selected at the contract level, not at the parent organization and market-area level. Thus, for the MA–VIP model, we mapped individual enrollee CAHPS surveys to a local market area using the enrollee’s county of residence and to a parent organization using the enrollee’s plan identifier. To be included in the model, a parent organization in a market area would have to have 100 or more surveys attributed to it. Based on the survey responses attributed to an MA–VIP reporting unit, we calculated case-mix-adjusted CAHPS measure results (e.g., getting needed care, rating of health plan) for each reporting unit.

Identify market areas and parent organizations to be included in the MA–VIP

Our MA–VIP model’s unit for assessing plan quality and payment adjustments is the MA parent organization in the local market area. Parent organizations are identified by CMS as reported by plans (e.g., United, Aetna, Kaiser Permanente, Anthem) and include national and regional plans. We use MedPAC market areas in our MA–VIP model.

Estimated number of market areas with sufficient parent organization enrollment to be included in the MA–VIP when implemented

To estimate how many local market areas would have sufficient parent organizations that meet enrollment requirements to calculate the illustrative MA–VIP measure set, we defined market areas as the roughly 1,200 MedPAC market areas designed to reflect local health care markets using 2017 MA plan enrollment data.¹⁸ To be included

in the model, each reporting unit (parent organization and market area) and peer group (where applicable) needed to meet the minimum sample size requirements identified in Table 3-6. For the HOS and CAHPS results, CMS would have to adapt MA–VIP requirements for fielding those surveys. An option is to apply a minimum sample of 600 to each reporting unit based on CMS’s current requirement that any contract with at least 600 enrollees must collect CAHPS results and a minimum of 500 enrollees must collect HOS results. Applying this requirement to each reporting unit (parent organization and market area combination) would likely increase the total number of surveys required, compared with the current number. However, MA plans currently field more than the minimum number of required surveys because they seek to oversample certain populations. Because the measure domains using HOS and CAHPS data do not use peer groups, the requirement to field the surveys would be 600 enrollees in each market area, regardless of full dual-eligibility status.

To determine the feasibility of applying the proposed reporting units to the MA–VIP model, we calculated the number of reporting units that meets the 600-enrollee requirement and the share of MA enrollment included in those reporting units.¹⁹ Table 3-7 (p. 76) shows the number of market areas with 0, 1, 2, and 3 or more parent organizations meeting the 600-enrollee requirement.

To implement the MA–VIP, we believe three parent organizations are necessary in a market area to ensure adequate comparison and distribution of rewards and penalties. Table 3-7 (p. 76) shows that there are 481 market areas with at least 3 parent organizations that meet a minimum sample of 600 enrollees; enrollment in those parent organizations accounts for 89 percent of MA enrollment. Our model’s minimum of 3 parent organizations excludes 749 market areas (generally nonmetropolitan areas), accounting for about 11 percent of MA enrollment, from the model’s bonus program, without a policy to include additional market areas.

Table 3-7 (p. 76) shows that 270 market areas have no parent organizations meeting the minimum enrollment of 600, covering 4 percent of current MA enrollment. It is not possible for the MA–VIP to operate in these markets. However, alternative approaches could be considered to include more of the 479 areas with only 1 or 2 parent organizations meeting the minimum enrollment of 600. One option is to combine market areas with too few parent organizations meeting the minimum criteria with

**TABLE
3-7****Illustration of the number of parent organizations in MedPAC market areas that met related MA enrollment requirements, 2017**

Category of market areas	Number of MedPAC market areas	Share of:		
		Total fully dual-eligible MA enrollment	All other MA enrollment	Total MA enrollment
Reporting units that met the 600-enrollee requirement for illustrative MA-VIP				
3 or more parent organizations	481	90.0%	88.6%	88.8%
2 parent organizations	233	3.5	5.5	5.3
1 parent organization	246	1.1	2.0	1.9
0 parent organizations	270	5.4	3.9	4.0
Total	1,230	100	100	100

Note: MA (Medicare Advantage). There are 1,230 MedPAC market areas designed to reflect health care markets. Parent organizations are the companies that operate the MA plans. We applied a minimum sample of 600 enrollees to the parent organizations based on CMS's current requirement that any plan with at least 600 enrollees must collect Consumer Assessment of Healthcare Providers and Systems® data and with at least 500 enrollees must collect Health Outcomes Survey results. Share of enrollment is based on the MA enrollment in the parent organizations in market areas that meet the category's criteria.

Source: MedPAC analysis of 2017 MA enrollment data.

contiguous market areas to meet the minimum three or more parent organizations with sufficient enrollment. For example, we could combine 499 Parent Organization A enrollees in X market area with the neighboring 11,200 Parent Organization B enrollees in Y market area for quality measurement and distributing MA-VIP rewards and penalties. Alternatively, the minimum number of parent organizations in a market area could be set at two. However, allowing only two parent organizations in a market area would consistently result in a direct transfer of dollars distributed from the worse performing parent organization to the better performing parent organization in that market area.

Number of market areas with sufficient quality results to be included in the MA-VIP model

Our model is also limited by the number of reporting units with sufficient data. Specifically, we are limited by the current availability of the HOS and CAHPS survey measures because the surveys are currently collected at the contract level and not at the parent organization and market-area level. Given this limitation, fewer parent organizations and market areas are included in our MA-VIP model.

After applying all criteria, our model includes 61 MedPAC market areas and 78 unique parent organizations for a total

of 258 reporting units (parent organization and market area combinations). On average, each market area includes about 4 parent organizations, ranging from 3 to 12 parent organizations in a market area. Using enrollment data from 2015 to 2017, these 61 areas represent about 40 percent of current MA enrollment (45 percent of fully dual-eligible enrollees and 39 percent of all other enrollees).

Convert performance on a small set of measures to MA-VIP points

Unlike the current QBP, which scores plans' performance on quality relative to other plans' performance scores, which are unknown until CMS applies the scoring, the MA-VIP is designed to reward or penalize a plan based on the plan's performance relative to prospectively set performance-to-points scales for each measure domain. For our MA-VIP model, we calculated each MA plan's performance (in the 61 market areas that met our criteria) on the 6 measures in the 4 measure domains we can include in the model. Using those results, we created a national performance-to-points scale for each measure. For the proposed MA-VIP model, we set the performance scale along a broad distribution of national historical data so that most plans would have the opportunity to earn points. We set the continuous points scale using a beta distribution, which helps to smooth the extremes of

**TABLE
3-8****Illustration of point system to score performance under our illustrative MA-VIP model****Measure domains and measures**

	ACS hospital use	Patient-reported outcomes		Patient/enrollee experience		Staying healthy and managing long-term conditions
	Risk-standardized rate of ACS hospitalizations per 1,000 MA enrollees (lower is better)	Improving or maintaining physical health status (higher is better)	Improving or maintaining mental health status (higher is better)	Getting needed care (higher is better)	Rating of health plan (higher is better)	Breast cancer screening (higher is better)
0 points	97	52%	73%	71%	76%	42%
2 points	59	61	80	81	82	63
4 points	49	64	82	84	84	69
6 points	41	67	84	87	86	74
8 points	33	70	86	89	88	80
10 points	16	77	90	95	92	92

Note: MA-VIP (Medicare Advantage value incentive program), ACS (ambulatory care-sensitive). Each of the six measures in the MA-VIP model is scored from 0 to 10 points; only a subset of points is displayed here. The national performance-to-points scale is based on the performance of the 258 parent organizations' peer groups in the 61 market areas with sufficient data to include in the model. The performance-to-points scales are set using a beta distribution. MA plans would technically receive only 10 points or 0 points with the best possible or worst possible scores, respectively.

Source: MedPAC analysis of MA quality data, 2015–2017.

a distribution by providing estimates of a true percentile independent of associated problems such as ceiling effects. MA plans earn up to 10 points for their performance on each of the 6 quality measures based on a continuous scale (see text box on reporting quality information, p. 59). Because Medicare is a national program, we score each plan's results against the same national performance-to-points scales, but a pool for dollars would be based on payments inside a market area and be given as rewards and penalties within that market area.

Table 3-8 presents a subset of our model's performance-to-points scale. The scales are set using historical data from the model's 258 reporting units. For example, the reporting unit in the lowest percentile of performance for the ACS hospitalization measure had a score of 97 ACS hospitalizations per 1,000 MA enrollees, which is set equal to 0 points. The reporting unit in the highest percentile of performance for that measure had a rate of 16 ACS hospitalizations per 1,000 MA enrollees, which

is set equal to 10 points. The MA-VIP scores each plan's peer group against the national standards. If a plan's peer-group ACS hospitalization score was about 33 ACS hospitalizations per 1,000 MA enrollees, it would earn about 8 points on that measure domain.

For each parent organization's peer group, we calculated a total MA-VIP score, which is a weighted average of the number of points earned for each domain. We followed CMS's QBP weighting approach with the most weight (factors of 3) given to the outcome domains (ACS hospital use and patient-reported outcomes), second highest weight (factor of 2) to the patient experience domain, and lowest weight (factor of 1) to the process measure (breast cancer screening).

Table 3-9 (p. 78) presents the average points across the peer groups in the 258 reporting units for the model's available measures. To convert performance to points for each peer group, we applied the national performance-to-points scale shown in Table 3-8 to each reporting unit.

**TABLE
3-9****Illustrative average MA-VIP points for each peer group****MA-VIP points, by measure domains and measures**

Peer group	ACS hospital use	Patient-reported outcomes		Patient/enrollee experience		Staying healthy and managing long-term conditions	Total MA-VIP
	Risk-standardized rate of ACS hospitalizations per 1,000 MA enrollees	Improving or maintaining physical health status	Improving or maintaining mental health status	Getting needed care	Rating of health plan	Breast cancer screening	
Fully dual-eligible enrollees (Peer Group 1)	3.5	5.0	5.1	5.1	4.9	3.5	46
All other enrollees (Peer Group 2)	6.4	5.0	5.1	5.1	4.9	6.6	54

Note: MA-VIP (Medicare Advantage value incentive program), ACS (ambulatory care-sensitive). Each measure in the MA-VIP model is continuously scored from 0 to 10 points. The national performance-to-points scale is based on the performance of the 258 parent organizations' peer groups in the 61 market areas with sufficient data to include in the model. The model scores two peer groups, one based on the quality of care provided to fully dual-eligible enrollees and one for all other enrollees.

Source: MedPAC analysis of MA quality data, 2015-2017.

The fully dual-eligible peer groups produced fewer points on average than the all-others peer groups on the ACS hospitalizations measure (3.5 points vs. 6.4 points) and on the BCS measure (3.5 points vs. 6.6 points), which is expected based on previous analyses of the BCS measure. The differential in the peer groups' performance supports the use of applying stratified payment adjustments to account for the social risk factors of plan populations. It is important that measure results not be adjusted because the differences in plan performance for the two populations could result from the problem that the fully dual-eligible population is more difficult to treat and manage or that plans covering those populations do not provide as high-quality care.

The patient-reported outcomes and patient/enrollee experience measures are survey based and are already case-mix adjusted for fully dual-eligible status, so a score based on a sample of the entire MA enrollment population is applied to both peer groups for MA-VIP scoring. On average, reporting units received 5.0 and 5.1 points for the patient-reported outcome measures and 5.1 and 4.9 on the patient/enrollee experience measures.

Converting MA-VIP points to payment adjustments using stratification into peer groups

Consistent with the Commission's principle that quality incentive programs should account for differences in providers' populations as needed, including social risk factors, our MA-VIP model stratifies the market-level populations it scores and redistributes pools of dollars for two peer groups: fully dual-eligible enrollees (Peer Group 1) and all other enrollees (Peer Group 2). The model uses eligibility for full Medicaid benefits, as in the HVIP, as a proxy for whether a plan's enrollees are more difficult to treat because these Medicare beneficiaries are much more likely than others to be disabled, have multiple chronic conditions, and have functional impairments. Policymakers should consider using other social risk factors to define peer groups, such as receiving the low-income drug subsidy, disability status (which is a current adjustment factor in the MA QBP), and area deprivation indexes, with the definitions subject to refinement as more data become available.

In each of the 61 market areas, for each of the 2 stratifications of enrollees, we created a pool of expected MA–VIP payments to plans based on 2 percent of each plan’s payments for its enrollees in that stratification. (As discussed earlier in the chapter, the percentage of plan payments that are used to create the pool of dollars could increase over time, and policymakers should consider the appropriate amount to incentivize quality improvement.) We also calculated the payment multiplier, or the percentage payment adjustment per MA–VIP point, which converts total points to dollars and results in spending each peer group’s pool of dollars. On this basis, we computed each plan’s MA–VIP payment adjustment by multiplying the peer group’s payment multiplier by the peer group’s total points earned.

In this way, the MA–VIP accounts for differences in social risk factors of plan populations and allows plans potentially to earn more rewards for higher quality care for their fully dual-eligible population than under the current QBP, owing to a higher payment multiplier for the fully dual-eligible enrollee peer group. This peer group on average has lower performance on quality measures, so when calculating a multiplier to redistribute a peer group’s pool of dollars, the multiplier will be higher than for the all-others peer group, which has higher MA–VIP points on average. In our MA–VIP model, we found that 93 percent of the market areas had higher percentage payment adjustments per quality point multipliers for the fully dual-eligible peer group.²⁰ These peer groups had a median payment multiplier of 0.42 percent (range: 0.30 percent to 0.74 percent); the all-others peer groups had a lower median payment multiplier of 0.35 percent (range: 0.23 percent to 0.50 percent). Thus, as intended, in the vast majority of market areas included in our model, plans have the potential to earn more points for high-quality care provided to their fully dual-eligible population.

Distribution of rewards and penalties by local market area and peer group

The MA–VIP will distribute rewards and penalties within each local market. However, for the Commission’s consideration, we produced results based on a national distribution of rewards and penalties (see text box illustrating a national distribution, pp. 82–83). For a discussion of the merits of local versus national distribution of rewards and penalties, see the section titled “Distribute rewards and penalties within local market areas” (p. 67).

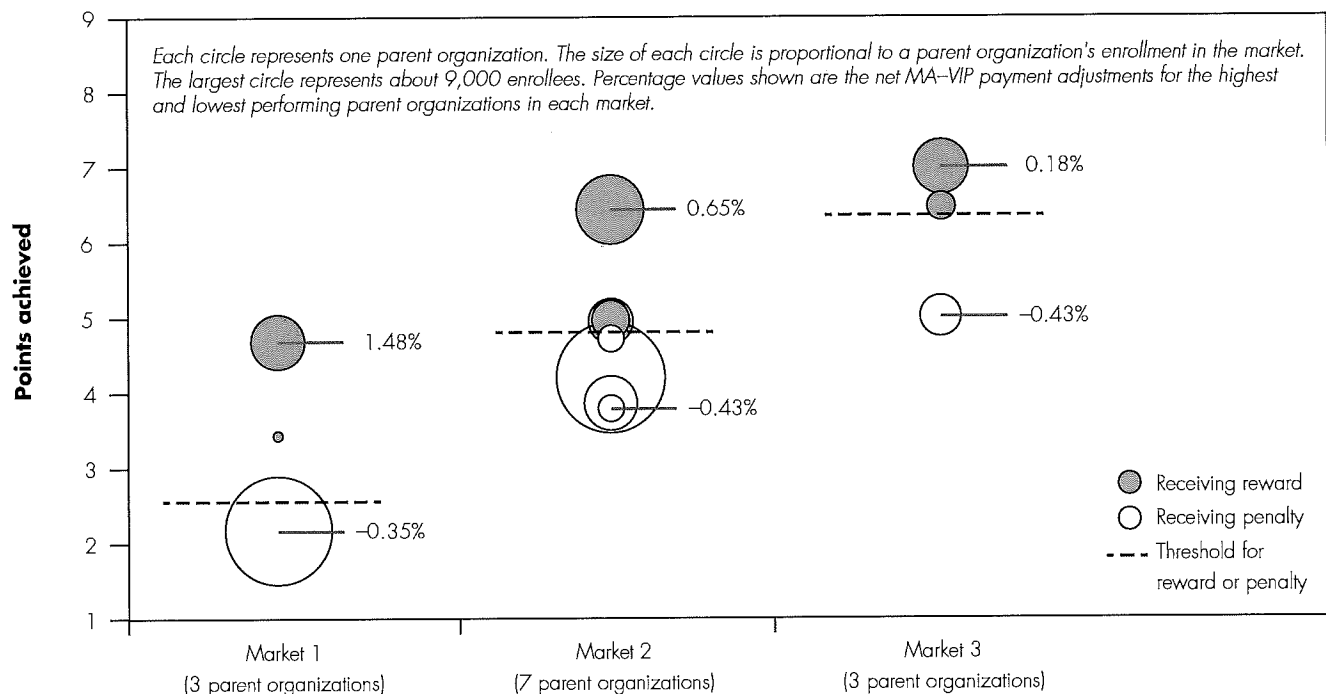
Under our MA–VIP model, a market area’s parent organizations with better quality scores (i.e., more MA–VIP points) receive a net positive payment adjustment, or a reward, and those with worse quality scores receive a net negative payment adjustment, or a penalty. Figure 3-5 (p. 80) and Figure 3-6 (p. 81) summarize the quality scores (MA–VIP points) achieved and net payment adjustments, by peer group, for the 78 parent organizations in the model’s 61 markets (totaling 258 parent organization and market observations for each peer group).

Figure 3-5 (p. 80) shows the results for three markets’ fully dual-eligible enrollees (Peer Group 1). Market 1 had low average performance (ranked 54th for both peer groups among all markets), Market 3 had high average performance (ranked 2nd and 3rd for each peer group), and Market 2 had average performance, near the middle of all markets. Parent organizations (shown with circles) in each market are distributed according to the average points achieved, and the size of each circle is proportional to enrollment. In Market 1, two parent organizations received a reward and one parent organization received a penalty. The dotted line in each market shows the threshold for receiving a penalty or reward in that market. Because rewards are distributed within each market, the threshold varies by market.

The size of any reward or penalty depends on the distribution of points achieved and distribution of enrollment among parent organizations in the market area. In Market 1, the parent organization with the largest share of fully dual-eligible enrollees in the market achieved 2.2 points and received a penalty of 0.35 percent, offsetting rewards of 0.57 percent (percent not shown) for a parent organization with very small enrollment achieving 3.5 points, and of 1.48 percent for a parent organization with moderate enrollment achieving 4.7 points. Figure 3-6 (p. 81) shows results for all other enrollees (Peer Group 2) for the same three markets.

Overall, parent organizations’ other-enrollee peer group (Peer Group 2) performed better—that is, scored more points under the model. Ninety-seven percent of parent organizations achieved more points for Peer Group 2 than Peer Group 1, and the thresholds for receiving a reward or penalty were higher in every market for Peer Group 2.

For both peer groups, the range of points across markets varied from about 1 to 5 points (out of 10 points) and was not strongly correlated with the average performance in the market (i.e., markets with higher average performance

**FIGURE
3-5****Illustrative MA-VIP: Points achieved and net payment adjustment for parent organizations in three markets, fully dual-eligible enrollees (Peer Group 1)**

Note: MA-VIP (Medicare Advantage value incentive program). Each circle represents one parent organization, where the size of the circle is proportional to enrollment for that peer group. Parent organizations receiving a reward are shown in gray, and parent organizations receiving a penalty are shown in white. The line in each market shows the threshold for receiving a penalty or reward in that market. The net MA-VIP payment adjustment (percent) is shown for the highest and lowest performing parent organizations in each market.

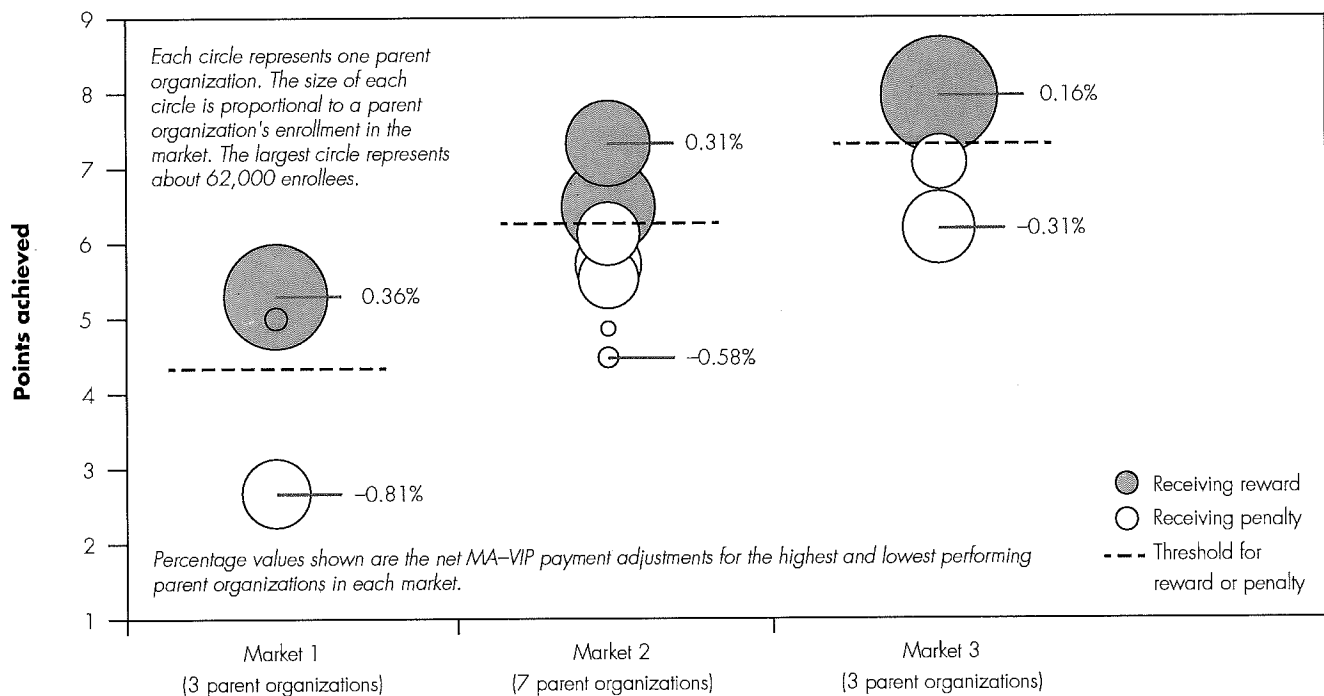
Source: MedPAC analysis of MA quality and payment data, 2015–2017.

did not tend to have a wider or narrower range of points achieved). However, the range of points achieved was moderately correlated with the number of parent organizations in the market, meaning that markets with more parent organizations tended to have a slightly wider range of points achieved.

Our MA-VIP modeling uses 2 percent of total plan payments as the basis for each reward pool; however, payment adjustments of nearly –2 percent or 2 percent would require extremes in performance in the same market. In our modeling, Figure 3-8 (p. 84) shows that net payment adjustments varied from a penalty of 1.5 percent of payment to a reward of 1.5 percent of payment for the fully dual-eligible enrollee peer groups (Peer Group 1), and from a penalty of 1.5 percent of payment to 1.0

percent of payment reward for the all-other peer groups (Peer Group 2). Most parent organizations in a market area had net payment adjustments between –0.5 percent (penalty) and 0.5 percent (reward) for each peer group.

Figure 3-9 (p. 85) shows the distribution of net payment adjustments aggregated to the parent organization (combining net payment adjustments across peer groups and market areas). Parent-organization payment adjustments ranged from about –1.1 percent to about 1.0 percent, with 76 of the 78 parent organizations receiving a net payment adjustment roughly between –0.6 percent and 0.6 percent, and a little more than half of all parent organizations receiving a net payment adjustment between –0.2 percent and 0.2 percent.

FIGURE 3-6**Illustrative MA-VIP model: Points achieved and net payment adjustment for parent organizations in three example markets, all other enrollees (Peer Group 2)**

Note: MA-VIP (Medicare Advantage value incentive program). Each circle represents one parent organization, where the size of the circle is proportional to enrollment for that peer group. Parent organizations receiving a reward are shown in gray, and parent organizations receiving a penalty are shown in white. The line in each market shows the threshold for receiving a penalty or reward in that market. The net payment adjustment (percent) is shown for the highest and lowest performing parent organizations in each market.

Source: MedPAC analysis of MA quality and payment data, 2015–2017.

The black bars in Figure 3-9 (p. 85) represent 11 parent organizations participating in 5 or more markets (and therefore tending to have greater total enrollment). Collectively, these parent organizations accounted for 161 of the observations in our model (62 percent) and received a reward 42 percent of the time. Because they participated in more markets and received both rewards and penalties, these parent organizations received offsetting rewards and penalties. The share of markets in which they received a reward ranged from 23 percent to 80 percent, whereas 55 parent organizations received only rewards or only penalties (45 of these parent organizations participated in only one market).

Modifying the magnitude of rewards and penalties

Because small rewards and penalties may not provide an adequate incentive for plans to improve quality, policymakers may want to increase the magnitude of rewards and penalties. Two aspects of the MA-VIP model could be modified to increase rewards and penalties: (1) the performance-to-points scale could be based on a truncated set of national results so that points achieved are disbursed more widely between 0 and 10 points, or (2) the size of the reward pools could be increased above 2 percent (possibly after a phase-in period) and based on a greater share of total payments. Either approach would

Illustration of national distribution of Medicare Advantage value incentive program rewards and penalties

An alternative approach to distributing rewards and penalties within each market would use a national distribution, whereby each peer group's reward pool would be distributed according to national performance results for all parent organizations in each market.²¹

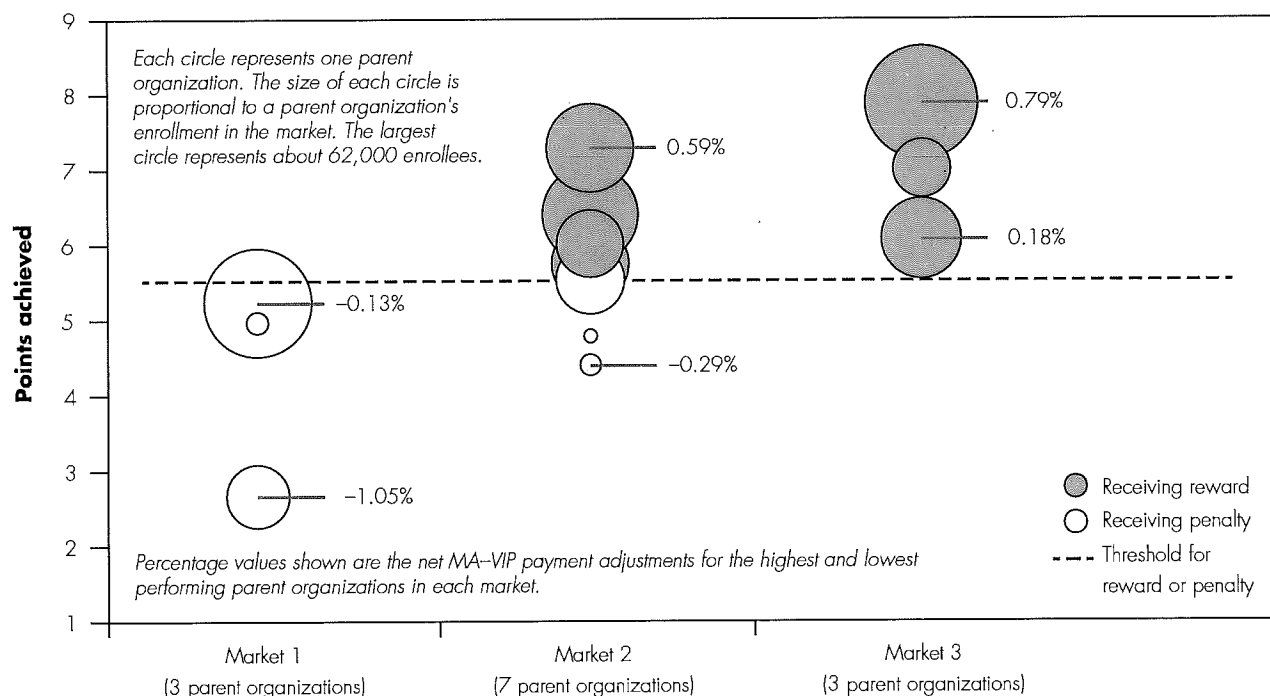
Figure 3-7 shows the results of a national distribution for the other-enrollees group (Peer Group 2). The threshold for receiving a national reward or penalty

in every market was about 5.6 points. Applying the national threshold causes all parent organizations in Market 1 to receive a penalty and all parent organizations in Market 3 to receive a reward. Under this approach for the 61 markets in our model, all parent organizations in 9 markets would have received a national penalty and all parent organizations in 8 markets would have received a national reward for Peer Group 2 (about 28 percent of markets were reward only or penalty only). About 79 percent of national rewards

(continued next page)

FIGURE 3-7

Illustration of national MA-VIP distribution of rewards and penalties: Average points achieved and net payment adjustment for parent organizations in three example markets, all other enrollees (Peer Group 2)



Note: MA-VIP (Medicare Advantage value incentive program). Each circle represents one parent organization, where the size of the circle is proportional to enrollment for that peer group. Parent organizations receiving a reward are shown in gray and parent organizations receiving a penalty are shown in white. The line in each market shows the threshold for receiving a penalty or reward in that market. The net MA-VIP payment adjustment (percent) is shown for the highest and lowest performing parent organizations in each market.

Source: MedPAC analysis of MA quality and payment data, 2015–2017.

Illustration of national distribution of Medicare Advantage value incentive program rewards and penalties (cont.)

for Peer Group 2 were distributed in the top half of markets (ranked by average market performance). These rewards were generated by penalties assessed mostly in the bottom half of markets, resulting in a broad transfer of rewards from lower performing markets to higher performing markets.

For fully dual-eligible enrollees (data not shown), the threshold for receiving a national reward or penalty was about 4.8 points nationally, and there were nine penalty-only markets and nine reward-only markets (about 30 percent of all markets). About 86 percent of all national rewards were distributed in the top half of markets (based on average market performance). ■

have the effect of stretching the distribution of net payment adjustments shown in Figure 3-8 (p. 84) and increasing the magnitude of rewards and penalties.²²

Comparison of MA-VIP model to existing MA QBP

Compared with the current QBP, our modeling demonstrates that the MA-VIP design would:

- address concerns about whether plans with large shares of high-needs populations are treated fairly,
- deal with geographic differences in the bonus status of population subgroups,
- eliminate the QBP's cliff effect, and
- not give an undue advantage to larger companies with more resources to manage the star system and companies that have benefited from contract consolidation.

Special populations

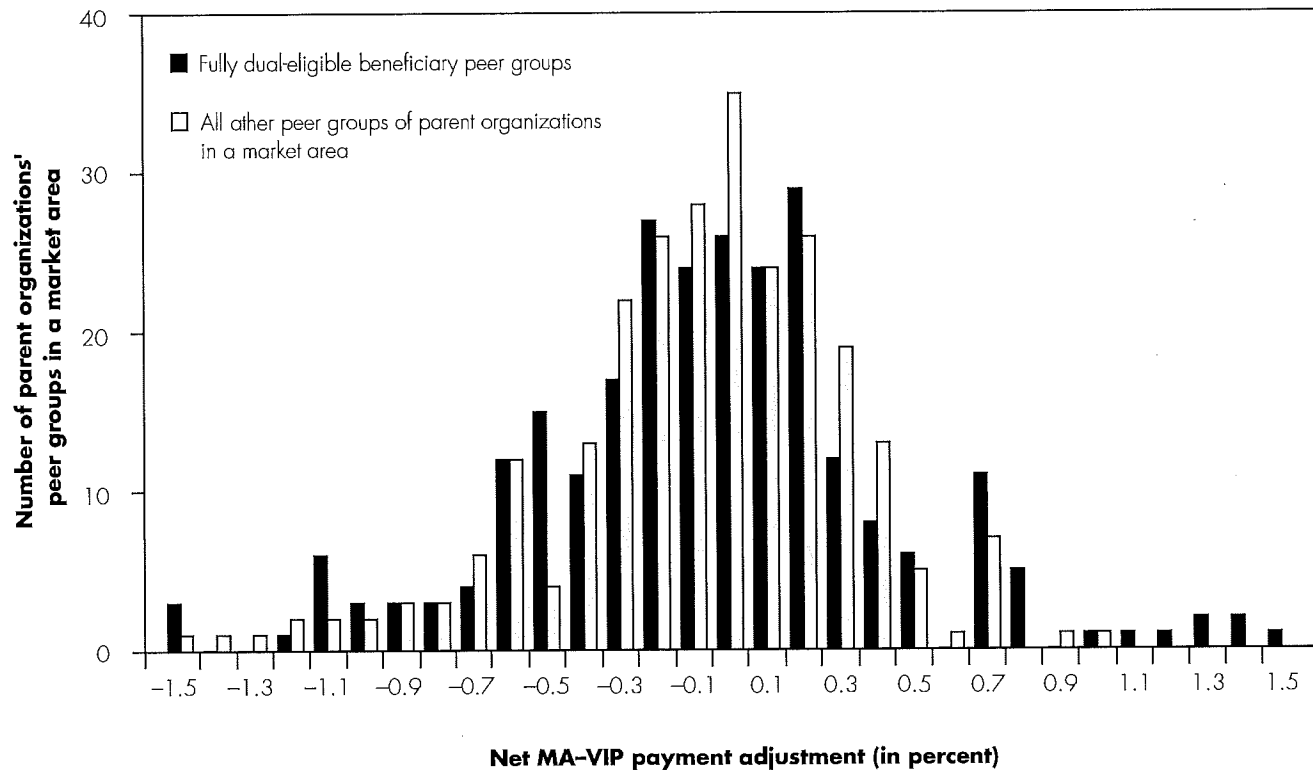
Under the QBP, with 83 percent of enrollees currently in bonus-level plans, nearly all MA enrollees are in plans deemed high quality. However, there are differences by population categories and by plan categories with respect to the rewarding of bonus payments under the QBP. Generally, plans with high shares of low-income enrollees, plans with high shares of enrollees under the age of 65 (entitled to Medicare on the basis of disability), and relatively smaller plans are less likely to have a bonus-level star rating. For example, in 2017, while about 75 percent of all MA enrollees were in bonus-level plans, the share among the fully dual-eligible population was

54 percent. For enrollees in employer group waiver plans (EGWPs, which provide MA coverage to employer-sponsored or union-sponsored retirees), the share in bonus plans was 92 percent in 2017. (EGWP status can be considered a proxy for higher income status, better health, and better access to health care.)

CMS employs a peer-grouping system in awarding star ratings so that plans with relatively higher shares of low-income beneficiaries and plans with higher shares of disabled beneficiaries have an adjustment to their star ratings—a feature intended to increase their likelihood of being in bonus status. However, the CMS peer grouping appears to only marginally change the bonus status of such plans. Our proposed MA-VIP instead uses a stratification approach to compare like populations. Under this model, an organization's performance for its fully dual-eligible population is compared with the performance of other organizations in the same market area for their fully dual-eligible population.

The MA-VIP stratification into peer groups and market-level comparison approach helps to level the playing field for plans serving fully dual-eligible beneficiaries (Figure 3-10, p. 86). Although in the QBP there are large differences in the share of fully dual-eligible beneficiaries versus other beneficiaries in bonus-level plans (54 percent vs. 82 percent in 2017), that difference is substantially narrower under the MA-VIP with respect to the share of enrollees in the MA-VIP peer group receiving positive financial results (53 percent vs. 57 percent).

The stratification used for the MA-VIP modeling separates only fully dual-eligible beneficiaries and all others. The Commission's past work has recognized

**FIGURE
3-8****Illustrative MA-VIP: Most peer groups of parent organizations in a market area receive small payment adjustments (rewards or penalties)**

Note: MA-VIP (Medicare Advantage value incentive program). The figure represents the distribution of net MA-VIP percent payment adjustments that peer groups receive after accounting for the 2 percent payment withhold used to create the pool of dollars to be redistributed. Changing the withhold percentage would expand or contract the distribution in line with the magnitude of the change in the percentage payment amount.

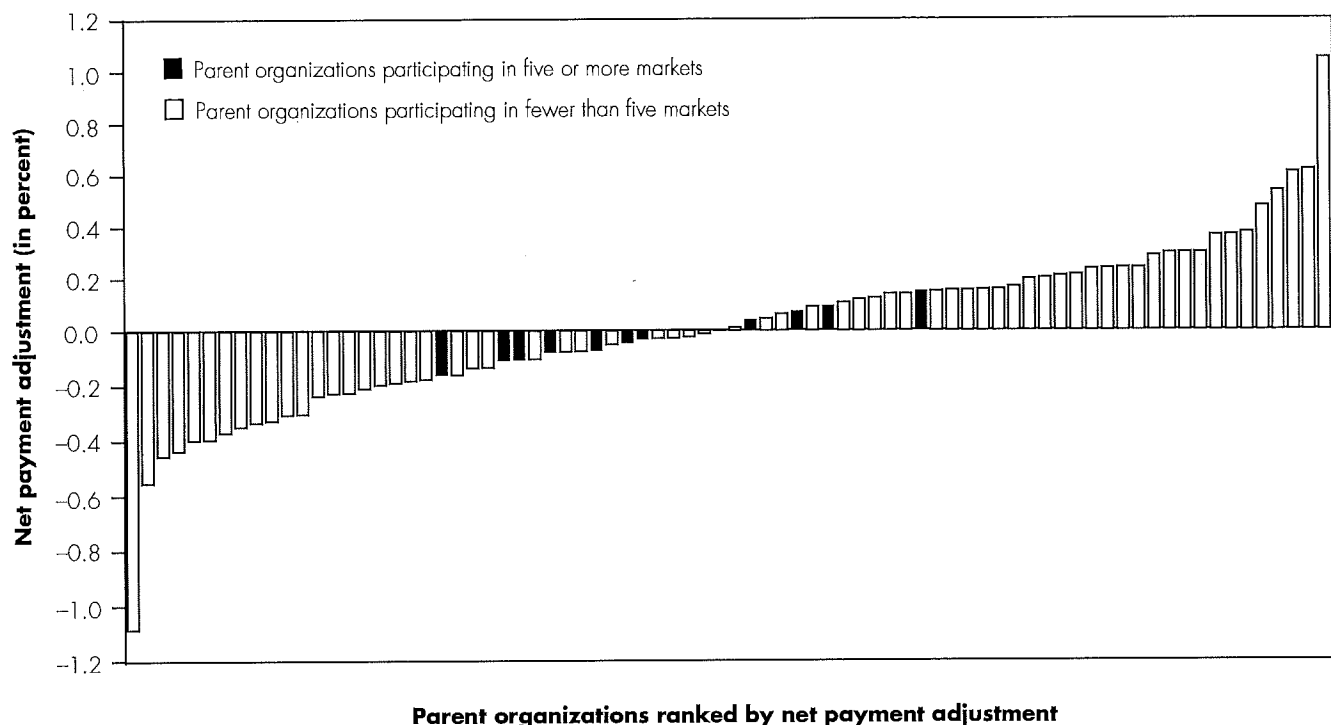
Source: MedPAC analysis of MA quality and payment data, 2015–2017.

systematic differences in quality results for the EGWP and under-65 populations, and CMS makes an adjustment to star ratings for contracts with high shares of enrollees originally entitled to Medicare on the basis of disability. Figure 3-10 (p. 86) shows that in the MA-VIP model, the EGWP population continues to fare better, and the under-65 population group fares slightly less well than the fully dual-eligible beneficiary group (51 percent for the under-65 population and 62 percent for EGWP enrollees, compared with 53 percent for the fully dual-eligible population). Our modeling is meant to be an illustrative prototype of how the MA-VIP could apply peer grouping, but when implementing the program,

it might be appropriate to stratify the EGWP and under-65 populations in addition to the fully dual-eligible population.

Eliminating the cliff and leveling size and sponsorship differences

With a continuous performance-to-points scale, and in part because of determinations made at the local market level, the MA-VIP design addresses another design flaw of the QBP system, in which plans lose bonus status if they fall short of a moving target that qualifies plans for bonuses (the “cliff” in the QBP). Only contracts with an average star rating of 3.75 (rounded to 4) or better on the

**FIGURE
3-9****Illustrative MA-VIP: Net payment adjustment by parent organization**

Note: MA-VIP (Medicare Advantage value incentive program). The figure represents the distribution of net payment adjustments received by parent organizations in the MA-VIP model. There are 78 distinct parent organizations in the illustrative MA-VIP model.

Source: MedPAC analysis of MA quality and payment data, 2015–2017.

5-star scale receive bonuses. In addition, in the QBP there are only small differences in the treatment of MA plans at or above 4 stars that could otherwise be distinguished because there are 4-star, 4.5-star, and 5-star contracts.

Among the plans included in our MA-VIP modeling, 20 parent organizations received no QBP bonus dollars in any of their markets. In our MA-VIP modeling, 8 of the 20 had positive results, ranging from a reward of 0.16 percent to 0.62 percent. These organizations are primarily regional plans (that is, plans operating in single markets or a small number of markets rather than organizations that have plans across the country). Six of the organizations operate in only one state, one operates in two states, and one has enrollment in five states.

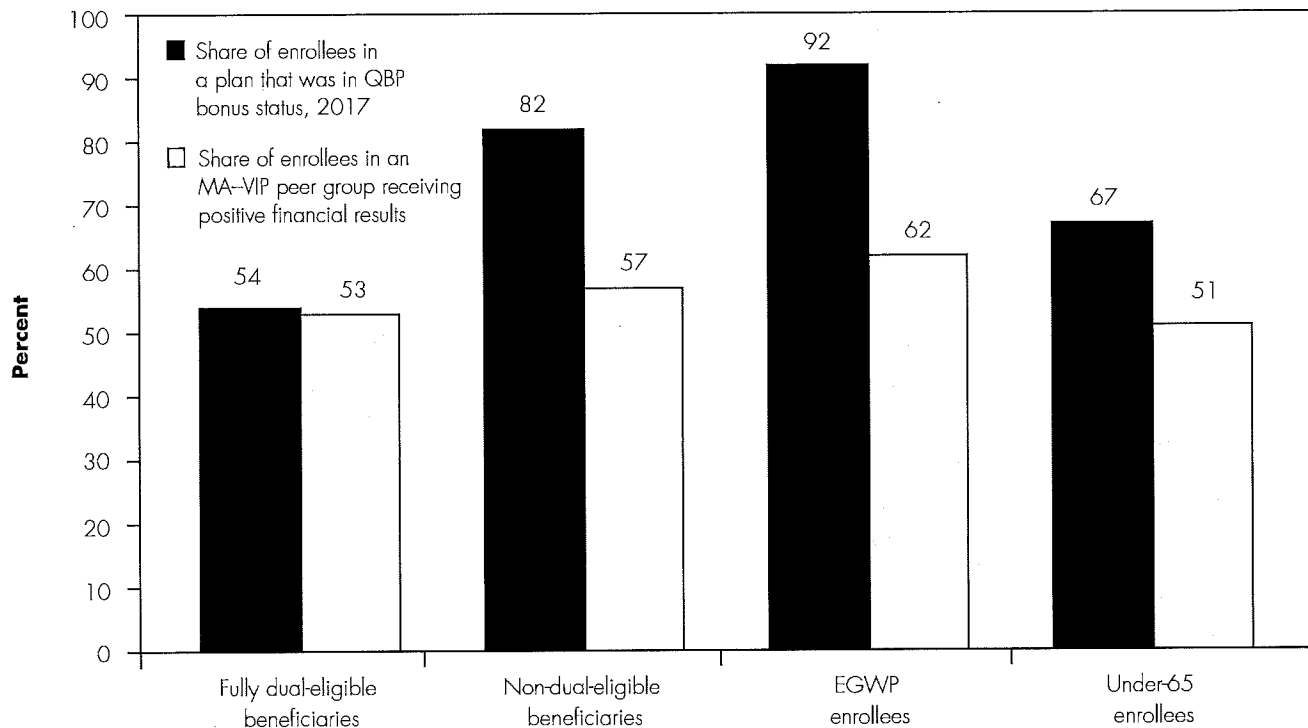
The eight organizations benefiting under the MA-VIP compared with their QBP status had relatively small

enrollment, illustrating that, under the QBP, differences exist between large contracts and small contracts. In the 2020 star ratings, 92 percent of enrollees in contracts with over 100,000 enrollees were in bonus-level contracts. For contracts with enrollment at or below 100,000, only 64 percent of enrollees were in contracts with bonus status. The larger contracts are often multistate contracts, and many are in bonus status as a result of contract consolidations. The three largest companies in MA enrollment have from 80 percent to 90 percent of their enrollment in bonus-level plans under the QBP. In the MA-VIP modeling, however, the performance of these organizations across markets varies significantly, and the companies have penalties in some markets and rewards in others.

A final point is that all the measures used in our model—other than the ambulatory care-sensitive hospitalization

**FIGURE
3-10**

For fully dual-eligible enrollees, plans have similar shares of enrollees in bonus status or positive illustrative MA-VIP results



Note: MA-VIP (Medicare Advantage value incentive program), QBP (quality bonus program), EGWP (employer group waiver plan). "Positive financial result" is an MA-VIP payment adjustment that exceeds the 2 percent of plan payments used to create the pool of dollars to be redistributed.

Source: MedPAC analysis of MA quality, enrollment, and payment data, 2015-2017.

measure—are used in the current star rating system to develop an average plan rating. Other measures include administrative measures (such as the timeliness of appeals processing), and some contracts have achieved their QBP bonus status on the strength of their good performance on measures other than outcome measures. Larger plans perform well on the administrative measures. If QBP stars were assigned solely based on the outcome measures used in our model, for 2020, only about one-third of MA enrollees would be in bonus-level plans, compared with 83 percent of enrollees currently in such plans. Rather than \$6 billion in added program payments, the added payments would be more in the range of \$2.3 billion.

Replacing the Medicare Advantage quality bonus program with a new value incentive program

Because of the many flaws of the QBP and the star system, the Commission asserts that Medicare lacks reliable information on which to evaluate quality within the MA sector. Fixing MA's quality measurement and quality incentive program is of the highest importance given that more than one-third of beneficiaries receive their care through MA plans and Medicare program expenditures for MA now total \$274 billion annually. The Commission has discussed moving Medicare into more value-based

payment models in which an entity is accountable for both the cost and quality of care provided to Medicare beneficiaries on a population basis. MA is such a model, but the current state of quality reporting and measurement in MA does not provide a basis for properly evaluating the effectiveness of this model, nor does the current system provide accurate information to beneficiaries. The flaws of MA quality measurement must be addressed so that Medicare can have confidence that high quality is being appropriately rewarded based on accurate information about plan performance.

Our exercise in calculating an illustrative MA–VIP prototype has demonstrated that it is feasible for the Medicare program to implement a system that addresses the QBP’s flaws. The model distributes both rewards and penalties to plans within market areas, based on plan performance on quality measures tied to clinical outcomes and patient experience. Under this model, most plans have the potential to receive higher rewards if their enrollee populations include large shares of enrollees with social risk factors. As compared with the QBP, the modeling results also show reduced disparity in plans’ financial performance with respect to fully dual-eligible enrollees compared with the financial performance for other enrollees.

The current practice of collecting data and measuring quality at the MA contract level limited the availability of data to use in our model; thus, the model is not meant to provide an exact formula for a QBP replacement. To make a program change, CMS should use the formal rule-making process to select measures, set performance-to-points targets, and define the social risk factors that are accounted for in peer groups.

RECOMMENDATION 3

The Congress should replace the current Medicare Advantage (MA) quality bonus program with a new MA value incentive program that:

- **scores a small set of population-based measures;**
- **evaluates quality at the local market level;**
- **uses a peer-grouping mechanism to account for differences in enrollees’ social risk factors;**
- **establishes a system for distributing rewards with no “cliff” effects; and**
- **distributes plan-financed rewards and penalties at a local market level.**

MA plans will be scored on their performance on quality and value measures, such as readmissions, patient

experience, patient-reported outcomes, and clinical care measures tied to outcomes. MA plan quality will be calculated at a local market level—for example, a parent organization within a market area instead of at the contract level. To account for differences in the social risk factors of plan populations, the MA–VIP will stratify results by defined peer groups, such as eligibility for Medicaid. Comparing groups with similar population characteristics accounts for social risk factors. We expect that as more data and research about the effects of patient-level social risk factors on quality performance become available, the approaches to assigning beneficiaries to a peer group will evolve.

The MA–VIP will reward or penalize a plan based on the plan’s performance relative to other plans in the market using a continuous, prospectively set performance-to-points scale for each measure. The MA–VIP redistributes a pool of dollars (made up of a percentage of plan payments within a market area) as rewards and penalties based on a plan’s performance compared with the market area’s other plans.

RATIONALE 3

The QBP is flawed and does not provide a reliable basis for evaluating MA quality in meaningful ways; plans have also received unwarranted bonus payments under the QBP system. Compared with the QBP, the MA–VIP will provide the program and Medicare beneficiaries with more accurate information on MA quality, and it is designed to produce a fairer distribution of incentive payments across markets and across the different population groups enrolled in MA.

The QBP currently costs the Medicare program \$6 billion a year in added program payments. Making the MA–VIP a plan-financed system that does not involve additional dollars will put the MA program on a par with nearly all FFS quality incentive programs, which are budget neutral or produce program savings. The Commission’s recommendation to replace the QBP with the MA–VIP produces program savings through reduced MA payments. The recommendation reflects the Commission’s interest in achieving equity in MA quality incentives and greater accuracy in determining plan eligibility for incentive payments. The recommendation is not intended as a strategy for establishing the appropriate level of overall payment to MA plans. In addition to developing an equitable system for quality-based payments, an assessment of overall payment adequacy for MA plans

should encompass all factors affecting MA plan payment, including policies for setting MA benchmarks and rebate levels, risk adjustment, and coding intensity—issues that the Commission has addressed, and will continue to address, in each year's March report to the Congress.

IMPLICATIONS 3

Spending

- This recommendation is expected to reduce program spending relative to current policy by more than \$2 billion over one year and by more than \$10 billion over five years.

Beneficiary and provider

- We do not expect this recommendation to have adverse effects on beneficiaries' access to plans or on plan participation in MA.
- It is possible that beneficiaries will see a reduction in extra benefits because plans will have lower payments; how much of a change there would be in extra benefits

depends on how plans respond to lower benchmarks and how they fare financially in the MA–VIP system. Bids could go up, but plans may also choose to reduce profits or otherwise lower their cost of providing the Medicare benefit—that is, they would become more efficient.

- To the extent that more money flows to plans serving high-needs populations, enrollees in those plans could have additional extra benefits. From the plan point of view, in addition to possible payment increases, the plans serving high-needs populations would be on a more even footing in competing with other plans in their area because of the stratification approach in determining rewards and penalties.
- With the MA–VIP, beneficiaries will have better information on the quality of plans in their area. Plans, however, will have higher administrative costs because of the use of the local area as the reporting unit. For example, more surveys will have to be administered. ■

Endnotes

- 1 Note also that, with respect to expected total expenditures for the QBP, the Congressional Budget Office estimate of a 10-year cost of \$94 billion if the QBP continues is an estimate that takes into account the Bipartisan Budget Act of 2018 provision requiring a weighted average of star ratings to determine the star rating of the surviving contract after a consolidation; that is, it takes into account the limited opportunities for future consolidation (Congressional Budget Office 2018).
- 2 HEDIS® is a registered trademark of the National Committee for Quality Assurance.
- 3 Almost all of the measure concepts in the illustrative measure set are part of the current MA star rating program and are included in the current Medicare ACO quality measure set.
- 4 Beginning with the 2021 star ratings, any changes to the measure set and scoring methodology will go through a formal rule-making process with notice and public comment. Before the 2021 star ratings, CMS announced and sought feedback on changes to the star ratings through the Part C and Part D call letter.
- 5 The relevant HEDIS measures currently available for plans to calculate using electronic clinical data systems include breast cancer screening and colorectal cancer screening. Plans can currently choose to report measure results through the traditional administrative data and medical record review or by incorporating data from electronic clinical data systems.
- 6 CAHPS® is a registered trademark of the Agency for Healthcare Research and Quality.
- 7 The HOS measures in the star system, and consequently in our modeling results, apply only to aged enrollees, even though enrollees under the age of 65 are also surveyed. CMS is considering using HOS results for the entire population of Medicare beneficiaries and has proposed expanding the minimum number of necessary responses from 30 to 100.
- 8 The CMS website includes files that identify the parent organization of each MA contract (<https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/MCRAdvPartDEnrolData/MA-Plan-Directory>). CMS defines a parent organization as “the legal entity that owns a controlling interest in a contracting organization...[and] is the ‘ultimate’ parent, or the top entity in a hierarchy (which may include other parent organizations) of subsidiary organizations which is not itself a subsidiary of any corporation. A legal entity may be its own parent organization if it is not a subsidiary of any other organization” (Centers for Medicare & Medicaid Services 2012).
- 9 In the hospital value incentive program, the term *peer group* means groups of hospitals. In the MA–VIP, *peer group* refers to groups of enrollees, sorted by various factors, including but not limited to social risk factors.
- 10 CMS also applies postmeasurement adjustments to overall star ratings, including a reward factor (rewarding a contract showing good performance across multiple measures), an improvement score for Part C and Part D improved performance that each has a weight of 5, and a “categorical adjustment index” that raises or lowers a contract’s overall star rating based on a contract’s share of low-income enrollees and the share of beneficiaries originally entitled to Medicare on the basis of disability (rather than age).
- 11 In prior years, outlier results for some measures distorted the clusters and inappropriately skewed the cut points identified by the clustering algorithm. For 2020, CMS modified the approach to put in place “guardrails” whereby, from one year to the next, the increase or decrease in cut points is limited to a 5 percent change (42 CFR §423.186(i)). CMS is also proposing to further reduce the effect of outliers (Centers for Medicare & Medicaid Services 2020).
- 12 A performance-to-points scale based on multiple years might simplify administration of the MA–VIP, but there is a tension between multiyear targets and the MA–VIP approach to financing. Revising targets each year would allow yearly calibration between (1) dollars expended as rewards or reduced payments through penalties and (2) the dollar amount that would most closely approximate budget neutrality in each year.
- 13 In the QBP, in addition to the incentive to achieve a 4-star rating and obtain bonuses, there are incentives to achieve an overall rating above 4 stars because contracts with a rating of 4.5 or 5 stars receive a higher level of rebate dollars, and 5-star plans can enroll beneficiaries outside of the annual election period. Policymakers will have to determine how these incentive provisions are treated in the MA–VIP system.
- 14 Plans apply administrative costs and profits to “load” the rebate dollars. The load averages 10 percent for extra benefits. When we report that in 2020 rebates are valued at \$122 per month, the “net” value to beneficiaries is about \$110 after accounting for the load. The \$27 figure includes the load, meaning that the net maximum change for beneficiaries would be \$24.
- 15 Comparison with FFS Medicare requires sufficient survey data within each market area. The CAHPS and the HOS are not fielded among FFS beneficiaries in each market, but

given sufficient funding, necessary survey data could be collected and available for comparison within a few years. MA encounter data have been found incomplete for some measures, and it is not clear when encounter data will be complete and available for all MA–VIP measures.

- 16 Hospitals are required to submit “no-pay” claims directly to CMS for all MA enrollees. Generally, these claims are a copy of claims hospitals submit to plans for payment. CMS uses no-pay claims in calculating disproportionate share hospital payments, medical education payments, and certain quality and utilization measures.
- 17 Using encounter data from 2015 to 2017, we calculated observed rates of ACS emergency department visits for the MA–VIP reporting units (i.e., parent organization within a market area) and found a distribution of visits suggesting that the outpatient encounter data are incomplete (including a number of reporting units with zero observed ED visits).
- 18 Metropolitan counties are grouped into a MedPAC market area if they are located in the same state and the same metropolitan statistical area. Nonmetropolitan counties are grouped into a MedPAC market area if they are located in the same state and the same health service area as defined by the National Center for Health Statistics. States can have multiple nonmetropolitan MedPAC market areas.
- 19 We used only the 600-enrollee criterion for this analysis because it is more limiting than the minimum sample size for ACS hospitalizations (150 enrollees) and about equally limiting as the readmission minimum sample size (150 admissions) using a rough average admission rate of 250 per 1,000 enrollees ($600 \text{ enrollees} \times 250 / 1,000 \text{ admission rate} = 150 \text{ admissions}$).
- 20 The market areas that did not have higher payment multipliers for the fully dual-eligible peer group had payment multipliers that were equal for both peer groups, or the all-others peer group payment multiplier was only a small percentage higher.
- 21 More specifically, to distribute the national share of the reward pool, the points achieved for each parent organization and each market for fully dual-eligible enrollees (Peer Group 1) would be pooled and a national pool of dollars would be distributed in one national market. The process would be repeated for the other-enrollees group (Peer Group 2).
- 22 In the MA–VIP design portion of this chapter, the section titled “Distribute plan-financed rewards and penalties at a local market level” (p. 62), we do not specify the share of plan payments that should be used to finance MA–VIP rewards. Policymakers should decide the appropriate level of plan payments to finance MA–VIP rewards.

References

Agency for Healthcare Quality and Research, Department of Health and Human Services. 2020. Talking quality: Reporting to consumers on health care quality. <https://www.ahrq.gov/talkingquality/index.html>.

Aligning Forces for Quality. 2009. *How to display comparative information that people can understand and use*. Research conducted on behalf of the Robert Wood Johnson Foundation. Washington, DC: Aligning Forces for Quality. <http://forces4quality.org/how-display-comparative-information-people-can-understand-and-use.html>.

Centers for Medicare & Medicaid Services, Department of Health and Human Services. 2020. Medicare and Medicaid programs; Contract year 2021 and 2022 policy and technical changes to the Medicare Advantage program, Medicare Prescription Drug benefit program, Medicaid program, Medicare Cost Plan program, and Programs of All-Inclusive Care for the Elderly. Noticed of proposed rulemaking. *Federal Register* 85, no. 32 (February 18): 9002–9260.

Centers for Medicare & Medicaid Services, Department of Health and Human Services. 2012. Memo to all current and prospective Medicare Advantage, prescription drug plan, cost and PACE organizations regarding verification of parent organization and legal entity name for 2013, March 2.

Congressional Budget Office. 2018. *Options for reducing the deficit: 2019 to 2028*. Washington, DC: CBO.

Government Accountability Office. 2012. *Medicare Advantage: Quality bonus payment demonstration undermined by high estimated costs and design shortcomings*. Washington, DC: GAO.

Marshall, M. N., P. G. Shekelle, H. T. Davies, et al. 2003. Public reporting on quality in the United States and the United Kingdom. *Health Affairs* 22, no. 3 (May–June): 134–148.

Medicare Payment Advisory Commission. 2019a. *A data book: Health care spending and the Medicare program*. Washington, DC: MedPAC.

Medicare Payment Advisory Commission. 2019b. *Report to the Congress: Medicare and the health care delivery system*. Washington, DC: MedPAC.

Medicare Payment Advisory Commission. 2019c. *Report to the Congress: Medicare payment policy*. Washington, DC: MedPAC.

Medicare Payment Advisory Commission. 2018a. *Report to the Congress: Medicare and the health care delivery system*. Washington, DC: MedPAC.

Medicare Payment Advisory Commission. 2018b. *Report to the Congress: Medicare payment policy*. Washington, DC: MedPAC.

Medicare Payment Advisory Commission. 2017. *Report to the Congress: Medicare payment policy*. Washington, DC: MedPAC.

Medicare Payment Advisory Commission. 2016. *Report to the Congress: Medicare and the health care delivery system*. Washington, DC: MedPAC.

Medicare Payment Advisory Commission. 2010a. *Report to the Congress: Aligning incentives in Medicare*. Washington, DC: MedPAC.

Medicare Payment Advisory Commission. 2010b. *Report to the Congress: Medicare payment policy*. Washington, DC: MedPAC.

Rose, A. J., E. Bayliss, L. Baseman, et al. 2019. Feasibility of distinguishing performance among provider groups using patient-reported outcome measures in older adults with multiple chronic conditions. *Medical Care* 57, no. 3 (March): 180–186.

Safran, D. G. 2019. Feasibility and value of patient-reported outcome measures for value-based payment. *Medical Care* 57, no. 3 (March): 177–179.

Wholey, D. R., J. B. Christianson, M. Finch, et al. 2003. Evaluating health plan quality 1: A conceptual model. *American Journal of Managed Care* 9 spec. no. 2 (June): SP53–64.

C H A P T E R

4

**Mandated report: Impact of
changes in the 21st Century
Cures Act to risk adjustment for
Medicare Advantage enrollees**

PFE000306

CHAPTER

4

Mandated report: Impact of changes in the 21st Century Cures Act to risk adjustment for Medicare Advantage enrollees

Chapter summary

The Medicare program pays managed care plans that participate in the Medicare Advantage (MA) program a monthly capitated amount to provide Medicare-covered services to its Medicare enrollees. Payment for each enrollee has two parts: a base rate and a risk score. The base rates vary by county, with the base rate for a given county reflecting the payment for an MA enrollee in that county with the health status of the national average beneficiary in fee-for-service (FFS) Medicare. The risk score indicates how costly the enrollee would be expected to be in FFS Medicare, relative to the national average FFS beneficiary.

The 21st Century Cures Act of 2016 directs the Secretary to make several changes to the CMS hierarchical condition category (CMS-HCC) model, which CMS uses to calculate enrollees' risk scores. The changes required by the 21st Century Cures Act include the following:

- Add indicators and adjustments for the total number of diseases or conditions for each enrollee.
- For beneficiaries who receive full Medicaid benefits, provide payment adjustments that are separate and different from payment adjustments for beneficiaries who receive partial Medicaid benefits. Until 2017, the CMS-HCC model had provided the same payment adjustment for these two beneficiary groups.

In this chapter

- Background on Medicare Advantage payments and risk adjustment
- Changes required by the 21st Century Cures Act
- Impacts of changes to CMS's risk adjustment model for Medicare Advantage
- Summary

- Evaluate the effects of including additional diagnosis codes for mental health disorders, substance abuse disorders, and chronic kidney disease.

In addition, the 21st Century Cures Act provides that the Secretary “may use at least two years of diagnosis data” to determine risk scores but does not appear to *require* the Secretary to do so.

CMS has implemented the mandated changes incrementally, applying different adjustments for full-benefit and partial-benefit dually eligible beneficiaries in 2017; adjustments for mental health disorders, substance abuse disorders, and chronic kidney disease in 2019; and adjustments for the number of conditions for each beneficiary in 2020. CMS has not implemented the use of two years of diagnosis data to determine risk scores.

The 21st Century Cures Act directs the Commission to evaluate the impact of these changes to the CMS–HCC model. To carry out this mandate, we evaluated five versions of the CMS–HCC model: the model that CMS used before implementing any of the changes mandated by the 21st Century Cures Act, the three models that CMS has implemented in response to the Act’s requirements, and a version that we developed that uses two years of diagnosis data to determine risk scores.

We evaluated how well each of the five versions of the CMS–HCC model predicts costs for various Medicare FFS beneficiary populations grouped by health characteristics, including type of medical conditions, number of medical conditions, and level of Medicare program spending. For each group, we calculated what each version of the CMS–HCC model predicts in costs for all of the group’s beneficiaries over one year (aggregate predicted costs). For each group, we also calculated how much Medicare actually spent on those FFS beneficiaries over one year (aggregate actual costs).

For each group, we determined a predictive ratio (PR), which is the ratio of aggregate predicted costs to aggregate actual costs. The desired result for a given group is a PR of 1.0, which would indicate that the model predicts costs for the group that are equal to the actual costs for the group. A PR less than 1.0 indicates that predicted costs for the group are less than actual costs, and the model will produce underpayments for that group. A PR greater than 1.0 indicates that predicted costs for the group are greater than actual costs, and the model will produce overpayments for that group.

In general, we found that:

- Each model produces accurate payment adjustments for groups that have characteristics defined by variables that are included in the model.
- Making distinctly different adjustments for full-benefit dual-eligible beneficiaries and partial-benefit dual-eligible beneficiaries eliminates systematic underpayments for the full-benefit dual-eligible beneficiaries and systematic overpayments for the partial-benefit dual-eligible beneficiaries that had occurred in previous models that did not distinguish between these two populations.
- Adding variables to the CMS–HCC model for mental health disorders, substance abuse disorders, and chronic kidney disease improves how accurately the model adjusts payments for beneficiaries who have those conditions. However, we caution that adding variables to the CMS–HCC model can provide additional opportunities for MA plans to increase revenue by coding more medical conditions. Such increases in coding may be especially likely when the additional variables represent conditions that are diagnosed using relatively discretionary standards.
- Adding indicators for the number of medical conditions for each beneficiary improves the model’s accuracy in adjusting payments for beneficiaries who have no conditions indicated in the model and those who have many conditions.
- All of the models produce underpayments for beneficiaries with very high levels of Medicare spending and overpayments for those with very low levels of Medicare spending. Adding indicators for the number of medical conditions for each beneficiary *slightly* improves the model’s accuracy in adjusting payments for both beneficiary groups, but underpayments and overpayments remain. These payment inaccuracies have been a persistent issue for MA risk adjustment.

We also found that using two years of diagnosis data to determine beneficiaries’ conditions produces payment adjustments that are about as accurate as using one year of diagnosis data, though it produces larger underpayments for those with high levels of Medicare spending than using one year of diagnosis data. Nevertheless, in our view, the use of two years of diagnosis data would be beneficial for MA risk adjustment because it would decrease the extent of coding differences that persist between the MA and FFS sectors of the Medicare program. Using two years of diagnosis data allows the model to capture more medical conditions among the FFS population, so that the profile of conditions among the FFS population more

closely matches the profile of conditions that would have been recorded for those beneficiaries had they been enrolled in MA. The result would be reduced payment errors that occur because of coding differences between MA and FFS.

We commend the progress that CMS has made in implementing the changes to the CMS–HCC model mandated by the 21st Century Cures Act. We encourage CMS to continue its work on this issue to complete the requirements in the 21st Century Cures Act by the mandated date of January 1, 2022. ■

Benefits of using prospective risk adjustment in Medicare Advantage

Two general arguments have been made for using a prospective risk adjustment model rather than a concurrent model in Medicare Advantage (MA):

- Relative to a concurrent model, a prospective model gives health plans greater incentive to manage their enrollees' care to prevent their enrollees from developing costly new conditions. Use of a concurrent model would move the MA program away from its intended purpose—managing the medical conditions of its enrollees—and closer to a cost-based model because plans would be paid as their enrollees' conditions occur.
- Plans face less uncertainty about their revenue streams under a prospective model. Under concurrent models, payments are based on conditions diagnosed in the prediction year. But it takes time for data on those diagnoses to be processed so that payments can be adjusted. Plans' revenue may then require adjustments after the prediction year ends. For example, if an MA enrollee had a condition diagnosed in December 2019, CMS may not be able to make an adjustment to the plan's payment until 2020. Under a prospective model, conditions from the base year are used to adjust payments in the prediction year, so the likelihood that payment adjustments are needed after the prediction year is smaller.¹ ■

Background on Medicare Advantage payments and risk adjustment

Medicare pays managed care plans that participate in the Medicare Advantage (MA) program a monthly capitated amount for each Medicare enrollee to provide Medicare-covered services. Each capitated payment has two general parts: a base rate, which reflects the payment for an MA enrollee with the health status of the national average beneficiary in fee-for-service (FFS) Medicare, and a risk score, which indicates how costly the enrollee is expected to be relative to the national average FFS beneficiary. The purpose of the risk scores is to adjust MA payments so that they accurately reflect how much an MA enrollee is expected to cost relative to the national average.

Over the years, CMS has used a variety of methods for determining MA enrollees' risk scores. Currently, CMS uses the CMS hierarchical condition category (CMS-HCC) risk adjustment model, which uses enrollees' demographic characteristics and medical conditions (such as diabetes and stroke) to predict their costliness. The demographic variables include age, sex, Medicaid status, institutional status, eligibility based on disability, and eligibility based on age but originally eligible because of disability.

CMS draws data for demographic variables from the year in which beneficiaries' costs are to be predicted (the prediction year). CMS bases assigned conditions on one year of diagnoses recorded on physician, hospital outpatient, and hospital inpatient claims from the year before the prediction year (base year). CMS groups the diagnoses into broader disease categories called hierarchical condition categories (HCCs). In the CMS-HCC model, some conditions have more than one HCC, which differ by severity of the condition. Examples include diabetes and cancer. The "hierarchical" part of HCC means that if a beneficiary has diagnoses that map into more than one HCC for a specific condition, only the HCC that has the largest effect on the beneficiary's risk score is used.

The CMS-HCC model is prospective, meaning it uses conditions from a base year to predict beneficiary costs in the next year (the prediction year), as opposed to concurrent, which uses conditions diagnosed in the prediction year to predict costs in the same year (see text box on prospective risk adjustment).

The purpose of risk adjustment is not to accurately predict costs for any particular person, but on average for a group of people with the same attributes that affect health care

costs. Therefore, an underlying feature of the CMS–HCC model is that, for beneficiaries who have the same HCC, it predicts costs that are below actual costs for some beneficiaries (underpredicts) and predicts costs that are above actual costs for others (overpredicts), but predicts accurately on average. This result is a feature of all models that use beneficiaries’ conditions to predict costs. This risk of loss faced by plans provides an incentive for plans to manage their enrollees’ conditions to keep their costs down. In addition, by paying accurately for each condition on average, the CMS–HCC model reduces incentives for plans to avoid beneficiaries with high-cost conditions.

Changes required by the 21st Century Cures Act

The CMS–HCC model is based on the standard HCC model developed by CMS (Pope et al. 2000). The CMS–HCC model differs from the standard HCC model in that it does not include all of the HCCs from the standard model. CMS has chosen not to use all of the HCCs because the agency believes that exclusion of some HCCs has a minimal effect on model performance while reducing burden on plans to submit data on their enrollees’ conditions and on CMS to process the data. However, by excluding some HCCs from the CMS–HCC model, CMS runs the risk of systematic underpayments to plans for enrollees with those conditions. In addition, CMS has always included in the CMS–HCC model an adjustment for whether a Medicare beneficiary receives some benefits from the Medicaid program (dual-eligible beneficiaries). Historically, CMS did not distinguish between dual-eligible beneficiaries with full Medicaid benefits from their state of residence and those with partial Medicaid benefits (their state paid their Medicare premiums and, in some cases, some of their Medicare cost-sharing responsibilities). However, the cost to the Medicare program is higher, on average, among the full-benefit dual-eligible beneficiaries relative to the partial-benefit dual-eligible beneficiaries. Consequently, risk adjustment that does not distinguish between these two populations produces systematic underpayments for full-benefit dual-eligible beneficiaries and systematic overpayments for partial-benefit dual-eligible beneficiaries.

In an effort to improve the CMS–HCC model, the Congress in the 21st Century Cures Act directed the

Secretary to make three changes to the CMS–HCC model (see text box on mandates, pp. 102–103):

- Add indicators for the total number of diseases or conditions for each enrollee.
- Provide separate payment adjustments for beneficiaries who receive full Medicaid benefits and for beneficiaries who receive partial Medicaid benefits. Until 2017, the CMS–HCC model had provided the same payment adjustment for these two beneficiary groups.
- Evaluate the effects of including additional diagnosis codes for mental health disorders, substance abuse disorders, and chronic kidney disease.

In addition, the 21st Century Cures Act indicates that the Secretary “may use at least two years of diagnosis data” to determine risk scores. It does not appear that use of two years of data is required.

CMS has implemented three of these changes indicated in the 21st Century Cures Act, the exception being use of at least two years of data to determine risk scores, when available.

The 21st Century Cures Act also directs the Commission to conduct an evaluation of the impact of these changes to the CMS–HCC model. In this report, we evaluated versions of the CMS–HCC model that CMS has implemented in response to the requirements in the 21st Century Cures Act:

- ***Different adjustments for MA enrollees with full Medicaid benefits and those with partial Medicaid benefits.*** The version of the CMS–HCC model that CMS used before 2017 (version 21, or V21) did not distinguish between these two groups of beneficiaries. In 2017, CMS implemented a version of the CMS–HCC model (V22) that distinguished between these two Medicare populations receiving Medicaid assistance by creating separate models for six population segments—
 - full Medicaid benefits and eligible for Medicare because of disability (disabled);
 - full Medicaid benefits and eligible for Medicare because of age (aged);
 - partial Medicaid benefits and disabled;

- partial Medicaid benefits and aged;
- no Medicaid benefits and disabled;
- no Medicaid benefits and aged.
- **Add HCCs for mental health disorders, substance abuse disorders, and chronic kidney disease.** For 2019, CMS implemented a version of the CMS–HCC model (V23) that added HCCs for mental health disorders, substance abuse disorders, and chronic kidney disease to V22 of the CMS–HCC model. CMS continued to use the six population segments from V22 in V23.
- **Include variables for the number of diseases or conditions for each beneficiary.** For 2020, CMS implemented a version of the CMS–HCC model (V24.1) that added indicators for the number of conditions for each beneficiary to V23. CMS determines the number of conditions for each beneficiary by counting the number of “payment HCCs” for each enrollee. A payment HCC is one that CMS includes in the CMS–HCC model used for payment purposes. CMS continued to use the six population segments in V22 and V23.
- **Use at least two years of diagnosis data to determine risk scores.** CMS has not implemented a version of the CMS–HCC model that uses two years of diagnosis data to determine risk scores. Nevertheless, we created and evaluated a version of the CMS–HCC model (V24.2) that is the same as model V24.1, but uses two years of diagnosis data. This version uses the same population segments used in models V22, V23, and V24.1.

We focused our evaluation of the changes that CMS has made to the CMS–HCC model on how well the resulting versions predict costs for Medicare populations defined by indicators of their health. The purpose of risk adjustment is to (1) adjust payments to MA plans such that those payments accurately reflect how much each MA enrollee is expected to cost in terms of covered services in FFS Medicare and (2) pay accurately enough so that plans do not have an incentive to attract beneficiaries because they would be profitable and avoid other beneficiaries because they would not be profitable. If risk adjustment does not pay accurately enough, plans could use beneficiaries’ health characteristics such as their medical conditions, number of health conditions, and historical health care costs to distinguish the favorable risks from the unfavorable risks.

In our analysis, we evaluated how well the CMS–HCC models implemented by CMS predict costs for beneficiary groups defined by health characteristics:

- Beneficiaries who have common medical conditions, including acute myocardial infarction (AMI), cancer, congestive heart failure (CHF), chronic obstructive pulmonary disease (COPD), mental illness, schizophrenia, all strokes, and ischemic or unspecified strokes.
- Beneficiaries stratified into groups by number of medical conditions, as indicated by the number of HCCs.
- Beneficiaries stratified by Medicare program spending in the year before the beneficiary’s risk score is determined (base-year spending). We determined the distribution of Medicare program spending among all Medicare beneficiaries and identified the percentile of each beneficiary’s Medicare spending. We then stratified the beneficiaries into these seven percentile categories: lowest 20 percent, 20 percent to 40 percent, 40 percent to 60 percent, 60 percent to 80 percent, 80 percent to 95 percent, 95 percent to 99 percent, and highest 1 percent.

We used predictive ratios (PRs) to evaluate how well the different versions of the CMS–HCC model predict costs for these various groups of Medicare beneficiaries. PRs indicate how well a model predicts costs for a group of beneficiaries with the same health characteristics. For a group of beneficiaries, a PR is the cost for the group as predicted by a risk adjustment model divided by the actual cost for that group. A PR greater than 1.0 indicates predicted costs are greater than actual costs for a group (overprediction); a PR less than 1.0 indicates predicted costs are less than actual costs for a group (underprediction). For a discussion of the details of our data and methods, see the text box about estimating and evaluating (pp. 104–105).

Impacts of changes to CMS’s risk adjustment model for Medicare Advantage

Our results indicate that each of the required changes CMS has made to the CMS–HCC model improves the predictive accuracy for each of the beneficiary populations that are the focus of the changes. Creating separate

Mandates to the Secretary and the Commission to improve risk adjustment in the 21st Century Cures Act

The legislative language of Section 17006 of the 21st Century Cures Act directs the Secretary and the Commission to improve risk adjustment in the Medicare Advantage program as follows:

(f) IMPROVEMENTS TO RISK ADJUSTMENT UNDER MEDICARE ADVANTAGE.—

(1) IN GENERAL.—Section 1853(a)(1) of the Social Security Act (42 U.S.C. 1395w–23(a)(1)) is amended—

(A) in subparagraph (C)(i), by striking “The Secretary” and inserting “Subject to subparagraph (I), the Secretary”; and

(B) by adding at the end the following new subparagraph:

“(I) IMPROVEMENTS TO RISK ADJUSTMENT FOR 2019 AND SUBSEQUENT YEARS.—

“(i) IN GENERAL.—In order to determine the appropriate adjustment for health status under subparagraph (C)(i), the following shall apply:

“(I) TAKING INTO ACCOUNT TOTAL NUMBER OF DISEASES OR CONDITIONS.—The Secretary shall take into account the total number of

diseases or conditions of an individual enrolled in an MA plan. The Secretary shall make an additional adjustment under such subparagraph as the number of diseases or conditions of an individual increases.

“(II) USING AT LEAST 2 YEARS OF DIAGNOSTIC DATA.—The Secretary may use at least 2 years of diagnosis data.

“(III) PROVIDING SEPARATE ADJUSTMENTS FOR DUAL ELIGIBLE INDIVIDUALS.—With respect to individuals who are dually eligible for benefits under this title and title XIX, the Secretary shall make separate adjustments for each of the following:

“(aa) Full-benefit dual eligible individuals (as defined in section 1935(c)(6)).

“(bb) Such individuals not described in item (aa).

“(IV) EVALUATION OF MENTAL HEALTH AND SUBSTANCE USE DISORDERS.—The Secretary shall evaluate the impact of including

(continued next page)

versions of the model for partial Medicaid beneficiaries and full Medicaid beneficiaries produces more accurate predictions of the cost of these beneficiaries. Further, adding indicators for mental health disorders, substance abuse disorders, and chronic kidney disease improves how well the CMS–HCC predicts the cost of beneficiaries who have these conditions. However, the addition of those indicators may increase opportunities for plans to boost revenue through more intensive coding. Finally, adding

measures of the number of conditions for each beneficiary improves how well the CMS–HCC model predicts the cost of beneficiaries who have 10 or more conditions.

We also found that using two years of diagnosis data to determine beneficiaries’ conditions produces payment adjustments that are about as accurate as using one year of diagnosis data, though this model produces larger underpayments for those with high levels of Medicare spending than using one year of diagnosis

Mandates to the Secretary and the Commission to improve risk adjustment in the 21st Century Cures Act (cont.)

additional diagnosis codes related to mental health and substance use disorders in the risk adjustment model.

“(V) EVALUATION OF CHRONIC KIDNEY DISEASE.—The Secretary shall evaluate the impact of including the severity of chronic kidney disease in the risk adjustment model.

“(ii) PHASED-IN IMPLEMENTATION.—The Secretary shall phase-in any changes to risk adjustment payment amounts under subparagraph (C)(i) under this subparagraph over a 3-year period, beginning with 2019, with such changes being fully implemented for 2022 and subsequent years.

“(iii) OPPORTUNITY FOR REVIEW AND PUBLIC COMMENT.—The Secretary shall provide an opportunity for review of the proposed changes to such risk adjustment payment amounts under this subparagraph and a public comment period of not less than 60 days before implementing such changes.”.

(2) STUDIES AND REPORTS.—

(A) REPORTS ON THE RISK ADJUSTMENT SYSTEM.—

(i) MEDPAC EVALUATION AND REPORT.—

(I) EVALUATION.—The Medicare Payment Advisory Commission shall conduct an evaluation of the impact of the provisions of, and amendments made by, this section on risk scores for enrollees in Medicare Advantage plans under part C of title XVIII of the Social Security Act and payments to Medicare Advantage plans under such part, including the impact of such provisions and amendments on the overall accuracy of risk scores under the Medicare Advantage program.

(II) REPORT.—Not later than July 1, 2020, the Medicare Payment Advisory Commission shall submit to Congress a report on the evaluation under subclause (I), together with recommendations for such legislation and administrative action as the Commission determines appropriate. ■

data. Nevertheless, in our view, the use of two years of diagnosis data would be beneficial for MA risk adjustment because it would decrease the extent of coding differences that persist between the MA and FFS sectors of the Medicare program. Using two years of diagnosis data allows the model to capture more medical conditions among the FFS population, so that the profile of conditions among the FFS population more closely matches the profile of conditions that would have been recorded for those beneficiaries had they been enrolled in MA. The result would be reduced payment errors that occur because of coding differences between MA and FFS.

Separate adjustments for fully dual beneficiaries and partially dual beneficiaries improves cost predictions

Since CMS began using the CMS–HCC model in 2004, CMS has included an adjustment for beneficiaries who are also eligible for Medicaid (Centers for Medicare & Medicaid Services 2003). However, being dually eligible does not mean all these enrollees have the same level of Medicaid coverage. Some have full benefits (FBs) from their state of residence, including prescription drugs, while others have only partial benefits (PBs), such as assistance with Medicare cost sharing and Medicare premiums. The

Method for estimating and evaluating versions of the CMS–HCC model

We used a sample of 27.2 million beneficiaries in fee-for-service (FFS) Medicare to evaluate five versions of the CMS hierarchical condition category (CMS–HCC) model, which CMS uses to risk adjust payments to Medicare Advantage (MA) plans. We randomly selected half this sample—13.6 million beneficiaries—to estimate coefficients in the five model versions:

- The version of the CMS–HCC model that CMS used in the MA program before 2017 (V21).
- The version of the CMS–HCC model that CMS used in the MA program in 2017 and 2018 (V22). This model is largely the same as V21, but CMS created separate adjustments for Medicare full-benefit dual-eligible beneficiaries (full Medicaid benefits from their state of residence) and for partial-benefit dual-eligible beneficiaries (their state pays their Medicare premiums plus cost sharing in some instances).
- The version of the CMS–HCC model that CMS used in the MA program in 2019 (V23). This model is largely V22, but CMS modified or added new hierarchical condition categories (HCCs) for moderate to severe substance abuse, minor substance abuse, reactive and unspecified psychosis, personality disorder, and Stage 3 chronic kidney disease.
- Model V24.1, which CMS began using in 2020, is V23 with additional categories for the number of conditions for each beneficiary. CMS defined the number of conditions as the number of HCCs that each beneficiary’s medical diagnoses map into.
- Model V24.2 (which is V24.1, but instead of using one year of diagnosis data to determine each beneficiary’s HCCs, V24.2 uses two years of diagnosis data when available). The Commission developed this model for this study.

We used the other half of the sample (13.6 million FFS beneficiaries) to evaluate model performance using predictive ratios (PRs), which indicate how well a model predicts costs for a group of beneficiaries

with the same health characteristics. For a group of beneficiaries, a PR is the cost for the group as predicted by a risk adjustment model divided by the actual cost for that group. A PR greater than 1.0 indicates predicted costs are greater than actual costs for a group (overprediction); a PR of less than 1.0 indicates predicted costs are less than actual costs for a group (underprediction). For this analysis, the prediction year is 2017, which is the year for which we are predicting beneficiaries’ costs. The previous year (2016) is the base year from which we draw beneficiaries’ conditions to determine their HCCs, except for V24.2, which has two base years (2015 and 2016) because we used two years of diagnosis data to determine HCCs.

All beneficiaries in our sample had Part A and Part B coverage in FFS Medicare in every month of 2016 (the sample for model V24.2 had Part A and Part B coverage in every month of 2015 and 2016). Beneficiaries must have lived within the 50 U.S. states throughout 2016 and must not have had Medicare as a secondary payer in 2016. In 2017, beneficiaries must have been in FFS Medicare for at least one month, must not have had Medicare as a secondary payer, must not have had end-stage renal status, must have lived within the 50 U.S. states throughout their enrollment in FFS Medicare, and must not have received hospice care.

For each beneficiary, we determined the months in 2017 during which the beneficiary was in a long-term care facility (living in an institution) and the months during which they were not (living in the community). During each of the months in which a beneficiary was living in the community in 2017, we assigned beneficiaries to one of these population segments:

- Full Medicaid benefits and eligible for Medicare because of disability (FULL_BENEFIT_DISABLED)
- Full Medicaid benefits and eligible for Medicare because of age (FULL_BENEFIT_AGED)
- Partial Medicaid benefits and eligible for Medicare because of disability (PARTIAL_BENEFIT_DISABLED)

(continued next page)

Method for estimating and evaluating versions of the CMS–HCC model (cont.)

- Partial Medicaid benefits and eligible for Medicare because of age (PARTIAL_BENEFIT_AGED)
- No Medicaid benefits and eligible for Medicare because of disability (NONDUAL_DISABLED)
- No Medicaid benefits and eligible for Medicare because of age (NONDUAL_AGED)

We assigned beneficiaries living in an institution to their own population segment, regardless of Medicaid status. For V22, V23, V24.1, and V24.2, we estimated separate versions for the seven population segments (institutionalized, FULL_BENEFIT_DISABLED, FULL_BENEFIT_AGED, PARTIAL_BENEFIT_DISABLED, PARTIAL_BENEFIT_AGED, NONDUAL_DISABLED, and NONDUAL_AGED).

For each beneficiary in our sample, we created the following variables for the regressions we used to estimate the coefficients (which indicate the additional cost of a characteristic or condition):

- The 2017 costs to the Medicare program incurred while in FFS Medicare. For each beneficiary, we divided these costs into the months the beneficiary was in each of the seven population segments (institutionalized, FULL_BENEFIT_DISABLED, FULL_BENEFIT_AGED, PARTIAL_BENEFIT_DISABLED, PARTIAL_BENEFIT_AGED, NONDUAL_DISABLED, and NONDUAL_AGED). We annualized these costs by dividing them by the fraction of 2017 that the beneficiary was in each of these segments.

- 24 age/sex categories for 2017
- Two categories—one for male, one for female—indicating whether a beneficiary was eligible for Medicare in 2017 because of age but was originally eligible because of disability
- The HCCs for each version of the model
- Disease interaction terms created from beneficiaries' HCCs
- For the institutional population, disabled/disease interaction terms

In estimating each model's coefficients, we used the beneficiaries' annualized 2017 FFS costs as the dependent variable and the remaining variables listed above as the explanatory variables in a weighted least squares regression. The weights were the fraction of 2017 that each beneficiary was in each of the seven population segments.

After estimating coefficients for V21, V22, V23, V24.1, and V24.2, we evaluated their efficacy using half of the 27.2 million-person full sample that we did not use to estimate the models. For each beneficiary, we determined the 2017 Medicare costs predicted by each of the five versions. We used these 2017 predicted costs to calculate PRs in nine disease categories, five HCCs that CMS introduced or modified for V23, seven categories that represent levels of beneficiaries' FFS costs in the base year of 2016, and nine categories that represent the number of conditions for beneficiaries in 2017. ■

versions of the CMS–HCC model that CMS used before 2017 did not distinguish between the FB beneficiaries and the PB beneficiaries, adjusting the capitated payments to MA plans by the same rate for all dual-eligible beneficiaries.

Research indicates that when the CMS–HCC model does not distinguish between FB beneficiaries and PB beneficiaries, the model systematically underpredicts

the cost for FB beneficiaries and overpredicts the cost for PB beneficiaries. For this analysis, we estimated how well CMS–HCC V21—which does not distinguish FB beneficiaries from PB beneficiaries—predicts costs for those two groups. Our analysis estimates an underprediction of 5 percent for FB beneficiaries and an overprediction of 5 percent for PB beneficiaries. In response to these systematic payment inaccuracies, CMS

made substantial changes to the CMS–HCC model for 2017. CMS replaced the single model for all enrollees that CMS identifies as living in the community (V21) by separating beneficiaries living in the community into population segments defined by their Medicaid eligibility status and their reason for Medicare eligibility (aged or disabled):

- full Medicaid benefits and eligible for Medicare because of disability (FULL_BENEFIT_DISABLED)
- full Medicaid benefits and eligible for Medicare because of age (FULL_BENEFIT_AGED)
- partial Medicaid benefits and eligible for Medicare because of disability (PARTIAL_BENEFIT_DISABLED)
- partial Medicaid benefits and eligible for Medicare because of age (PARTIAL_BENEFIT_AGED)
- no Medicaid benefits and eligible for Medicare because of disability (NONDUAL_DISABLED)
- no Medicaid benefits and eligible for Medicare because of age (NONDUAL_AGED)

CMS has also maintained a distinct version of the CMS–HCC model for enrollees who lived in an institutional facility (primarily nursing homes) for at least three consecutive months in the prediction year. Therefore, we evaluated how well risk adjustment predicts costs for seven population segments: six population segments in the community that are distinguished by their Medicaid status and whether they are Medicare eligible because of age or disability and one population segment for the long-term institutionalized.

We determined PRs for each of the versions for the six community population segments and for the institutional population. For all seven population segments, we found that V22 produced PRs of 1.00 for the entire population in the model.² These results indicate that the model pays accurately for each of the population segments, on average. In other words, separately estimating the model for each of the six population segments results in accurate payments for both beneficiaries who have full Medicaid benefits and beneficiaries who have partial Medicaid benefits (as well as those who have no Medicaid benefits). These accurate payments for population segments are an improvement over the single version of the CMS–

HCC model that CMS used for the community and institutionalized populations before 2017 (V21).

We also evaluated how well CMS–HCC model V22 predicts costs for groups of beneficiaries defined by other health characteristics. For each of the six community population segments and the institutional population, we:

- grouped beneficiaries by several medical conditions in the base year (AMI, cancer, CHF, COPD, diabetes, mental illness, schizophrenia, all stroke, and ischemic or unspecified stroke);
- stratified beneficiaries by the number of medical conditions, as indicated by the number of HCCs;
- stratified beneficiaries by their cost to the Medicare program in the base year of 2016; and
- grouped beneficiaries by whether they had one of the HCCs that CMS added to or modified for the CMS–HCC model in 2019 (moderate to severe substance abuse, mild substance abuse, reactive and unspecified psychosis, personality disorder, and Stage 3 chronic kidney disease).

We chose these health characteristics because they can be observed by plans, and, therefore, plans can use these characteristics to select enrollees. Plans cannot use other characteristics such as beneficiaries' cost to the Medicare program in the prediction year (2017 for this study) because the plans cannot observe these characteristics before beneficiaries make their decisions about MA enrollment.

We found that V22 predicts accurately in each population segment for the conditions included in the model (Table 4-1, p. 108). The greatest degree of underprediction is for schizophrenia among the FULL_BENEFIT_AGED segment (PR = 0.97), and the greatest degree of overprediction is for AMI among the FULL_BENEFIT_DISABLED and NONDUAL_AGED segments (PR = 1.02). However, neither result indicates a large payment inaccuracy.

When we stratified beneficiaries by the number of conditions they had (which is not a variable in V22 but which CMS added for V24.1), we found that for each of the seven population segments, V22 predicted well for beneficiaries who had from one condition to eight conditions. However, we found some degree of underprediction in all population segments for beneficiaries with no conditions indicated in the model